

Adversarial inference: predictive minds in the attention economy

Jelle Bruineberg^{*,†}

Center for Subjectivity Research, Department of Communication, University of Copenhagen, Karen Blixens Plads 8, Copenhagen 2300, Denmark; Department of Philosophy, Macquarie University, Balaclava Rd, Macquarie Park, New South Wales 2109, Australia

†Jelle Bruineberg, <http://orcid.org/0000-0001-5675-3376>

*Corresponding author. Center for Subjectivity Research, Department of Communication, University of Copenhagen, Karen Blixens Plads 8, Copenhagen 2300, Denmark. E-mail: bruineberg@hum.ku.dk

Abstract

What is it about our current digital technologies that seemingly makes it difficult for users to attend to what matters to them? According to the dominant narrative in the literature on the “attention economy,” a user’s lack of attention is due to the large amounts of information available in their everyday environments. I will argue that information-abundance fails to account for some of the central manifestations of distraction, such as sudden urges to check a particular information-source in the absence of perceptual information. I will use active inference, and in particular models of action selection based on the minimization of expected free energy, to develop an alternative answer to the question about what makes it difficult to attend. Besides obvious adversarial forms of inference, in which algorithms build up models of users in order to keep them scrolling, I will show that active inference provides the tools to identify a number of problematic structural features of current digital technologies: they contain limitless sources of novelty, they can be navigated by very simple and effortless motor movements, and they offer their action possibilities everywhere and anytime independent of place or context. Moreover, recent models of motivated control show an intricate interplay between motivation and control that can explain sudden transitions in motivational state and the consequent alteration of the salience of actions. I conclude, therefore, that the challenges users encounter when engaging with digital technologies are less about information overload or inviting content, but more about the continuous availability of easily available possibilities for action.

Keywords: attention economy; attention; active inference; distraction; digital technology

Introduction

In the competition for attention, digital platforms build up models of their users: where they click, what keeps them watching, what triggers them. These models are used to present the user with content and structure that optimizes the platform’s “engagement goals.” This much is the central premise of the “attention economy,” the business model in which human attention is the scarce resource over which digital platforms compete for their existence. Needless to say, there is increasing concern about the effects of living in the attention economy on the user’s mental lives, their agency, and their autonomy (Williams 2018; Zuboff 2019; Castro and Pham 2020; Chomanski 2023).

There is a striking parallel between the functional description of the algorithms involved in the attention economy and predictive-processing theories of the human mind. According to predictive processing, adaptive behavior is achieved by an agent building up a model of its environment. This model, in turn, is used to generate predictions about which actions will most likely lead to outcomes that are in line with the agent’s desired observations. Through the continuous adaptation of perception and action (a

process known as “active inference”), an agent learns to exploit the regularities in its environment to ensure favorable observations (Parr et al. 2022).

In this paper, I will explore the interplay between the agent as predicting its exchanges with the environment (central to active inference) and the (smart) environment as actively predicting its exchanges with the user. By further developing this interplay, we can better understand what it is about digital technologies that make them demand engagement, and, consequently, to better understand the vulnerabilities of the embodied mind situated in an increasingly designed and digital environment.

In popular treatments on the attention economy, such as in the documentary *The Social Dilemma* (Orlowski 2020), the user is depicted as a passive marionette, at the mercy of whoever is able to pull its strings. As we will see later in the paper, active inference paints a much more active picture of the mind in which action selection is dominated by both a pursuit of desired observations and a pursuit of uncertainty-reducing observations (especially in places where it expects uncertainty to be found). One central claim I will develop in this paper is that the challenges users encounter are less about seductive content (i.e. clickbait), and

Received 28 February 2023; Revised 18 July 2023; Accepted 31 July 2023

© The Author(s) 2023. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

more about the continuous availability of easily available opportunities for uncertainty-reduction. The expectation of uncertainty-reduction gives rise to the urges in the user to regularly check out the opportunities for novelty (Oulasvirta et al. 2012).

I will start the paper by putting some more substance to the concept of the “attention economy” and its underlying conception of attention. Second, I will situate the conception of attention within the debate in the psychology of attention about attention being a “cause” or an “effect.” As an intermediate finding, I will conclude that the conception of attention, as used in the debate on the attention economy, is problematic and that selection for action models are better able to capture the relevant phenomena. I will then introduce active inference and motivate why I think active inference can provide a better understanding of attention and the way it is dependent on the structure of the environment. I will focus on three distinct aspects of active inference in the context of the attention economy: the idea that both agent and environment are mutually predicting one another. Second, the specific conceptualization of attention, salience, and action in contemporary accounts of active inference. And third, the relatively underexplored idea that effort biases action selection in active inference.

The attention economy

According to Wu (2017), the attention economy started with the first newspaper beginning to use advertisement, rather than paid subscriptions, as its primary source of income. In contemporary times, the attention economy refers to the idea that the majority of digital platforms exist by virtue of their capacity to attract large amounts of “attention.” The phenomenon of the “influencer” shows that the business model of the attention economy is not limited to tech companies, but also pertains to competition between its users (academics included). According to the logic of the attention economy, the present text has “won” in the competition with all the other things that you, dear reader, could have been doing right now.

Following this logic, a consequence of the attention economy is that players in the attention economy (be it free newspapers, social media platforms, or influencers) start to tailor the content they produce to what draws most attention (Nguyen 2021). The business model pushes for “a race to the bottom of the brainstem” (Harris 2019) in which companies are ever better able to exploit the vulnerabilities of its users. The result is “an assault on human autonomy” (Zuboff 2019), in which the users’ attention is not aligned with their supposedly noble personal level interests, but with their most “primordial desires” (sex, violence, outrage, gossip, etc.). Perhaps even more disconcerting, it seems like one way of keeping a user engaged is to present them with more and more radical content, such that eventually their personal level interests threaten to shift in a more radical and extremist direction (Alfano et al. 2021).

While this might be one aspect of what it is to be a “user” in the attention economy, it cannot be the whole story. A second consequence of the attention economy is said to be increased distraction: we do not just pay attention to different things than we would like to, there is a sense in which we do not pay attention at all. In this second sense of attention, more related with notions as focus and concentration, our mental lives are becoming more fragmented overall: “the mind is simply jumping about and following *whatever grabs it*” (Wu 2017, p. 180, my italics). While writing this text, I may suddenly find myself scrolling through my Twitter feed without really having noticed the transition. In many of such instances, there is no prompt or notification that guides the user

to transition to Twitter. The transition seems to be self-initiated, but without anything like volition being involved.

In his guidebook on how to hook users to one’s product, Eyal (2014) distinguishes between “external” and “internal” triggers. External triggers are salient stimuli in the environment that attract attention. Consider, for example, the attention-grabbiness of the red circle with a white number that many apps use to indicate a user has unread messages. The “action” that is “triggered” by the prompt is to click on the app and see what new messages there are. Focusing on external triggers leads to the attention economy as it is commonly portrayed: an environment full of salient stimuli that compete for the users’ attention/engagement. In some cases, the salience of a stimulus seems to derive from indeterminacy: the user does not know what e-mails await behind the notifications. The urge to click and open the app seems primarily epistemic. In other cases, the salience seems to derive from a particular mix of hedonic and epistemic factors as, for example, in the case of “clickbait.” (Wu 2017, p. 282) describes clickbait as:

sensationally headlined articles, paired with provocative pictures - a bikini-clad celebrity was always good [...]. When properly calibrated, such content seemed to take control of the mind, causing the hand almost involuntarily to click on whatever was there.

The persuasive force of clickbait, and how it seems to sidestep conscious control, is itself a highly fascinating and interesting phenomenon that deserves further interest.¹ However, the attention economy is more than the competition of external triggers for attention and engagement. After all, external triggers presuppose that the user is already in perceptual contact with the notifications of a particular platform. If you are anything like me, you have switched off distracting notifications and pop-ups a long time ago. According to Eyal, the “brass ring of consumer technology” is therefore to intervene in the process one step before, by connecting “internal triggers with a product” (2014, p. 48). Internal triggers are prompts for action that arise in the absence of stimuli: you stand waiting for the train and feel an urge arising to check your e-mails. Or you are engaged in a conversation and feel an urgency to check your social media accounts (a phenomenon known as phubbing (Aagaard 2020)). Eyal writes that the aim here is to instill in the user an imbalance or stress “so that the user identifies the company’s product or service as the source of relief” (2014, p. 52).

On Eyal’s website, the model for how to instill internal triggers in a user is introduced under the heading *Hooks: An Intro on How to Manufacture Desire in 4 Steps* (Eyal 2012). Indeed, the competition for attention is an “epic struggle to get inside our heads” (Wu 2017, subtitle). Note though, that this is “desire” in scare quotes. These are not consciously endorsed desires or preferences, but desires embodied by sensorimotor engagement (perhaps more akin to conative “aliefs” (Gendler 2008)). I will return to this issue in a moment. For now it is important to point out the deep misalignment between the user’s own explicit goals and desires, and the ones “manufactured in them” through engagement with digital technologies. Williams writes: “No one wakes up in the morning and asks: ‘How much time can I possibly spend using social media today?’” (2018, p. 8). Still, these are the targets employees of tech companies set out to optimize in shaping our behavioral cycles.

¹ Eyal’s use of the term “trigger” suggests a reflex-like relation between salient stimulus and action. This literature is often ambiguous between the stimulus-response relation being a pure reflex, or as stimuli providing a highly persuasive demand for action. In this light, Wu’s “almost involuntarily” conveys deep ambiguity about the underlying philosophy of mind.

As Jeff Hammerbacher puts it succinctly: “The best minds of my generation are thinking about how to make people click ads. That sucks.” (as quoted in [Wu 2017](#), p. 30).

Zooming in on attention

Despite its terminological prominence, the concept of attention itself has been poorly theorized in the literature on the attention economy. Prominent authors like [Williams \(2018\)](#) and [Pedersen et al. \(2021\)](#) point to Herbert Simon’s essay *Designing Organizations For An Information-Rich World* ([Simon 1971](#)) for providing the initial formulation of what we now know as the attention economy.² According to Simon’s analysis, humans have long lived in a world of “information scarcity.” With the rise of new technologies, mass media, and communication, information has become excessively available. Consequently, we now live in an “information-rich world”:

In an information-rich world, the wealth of information means a dearth of something else: a scarcity of whatever it is that information consumes. What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it. ([Simon 1971](#), pp. 40-41)

On Simon’s account, information and attention are inversely proportional quantities: information-abundance implies attention-scarcity and vice versa. He follows the logic of resource allocation as used in economics: the wealth of information in our society creates the problem of “allocating” the scarce resource of attention to the abundance of information sources that can consume it.

In his 1971 paper, Simon’s concern was not with consumers bombarded by stimuli fine-tuned to draw their attention away. Instead, his paper deals with designing information-flows in large organizations to avoid bottlenecks. His focus lies on bureaucrats that receive more reports than they have time to read, and hence cannot process and pass information on efficiently. Throughout the paper, Simon assumes that the agent themselves allocates their attention to particular (attention-consuming) tasks. His analysis of the attention economy therefore does not seem to have much space for distraction, triggers, and lack of focus in the sense discussed above.

As far as I am aware, it is only James Williams who draws the connection between information abundance, in Simon’s sense, and a loss of control or agency. Williams writes:

[T]he main risk information abundance poses is not that one’s attention will be *occupied* or *used up* by information, as though it were some finite, quantifiable resource, but rather that one will *lose control over* one’s attentional processes. . ([Williams 2018](#), p. 15)

To illustrate his point, [Williams \(2018\)](#) compares reaching the informational threshold to progressing in a game of Tetris. As more and more chunks of information come in at an increasingly rapid pace, we lose our capacity to process and store the chunks in the way we want. What we need to do then, following this logic, is to

reduce the amount of information in our environment to come to regain control over our attention:

To say that information abundance produces attention scarcity means that the problems we encounter are now less about breaking down barriers between us and information, and more about putting barriers in place. ([Williams 2018](#), p. 16)

There is no doubt that putting barriers in place between the user and information is part of the solution. Indeed, switching off notifications on one’s phone is something like putting a barrier in place between the user and (irrelevant) information. But the interesting question is whether the barrier serves to keep the outside world from coming in, as Simon and Williams argue, or the inside from breaking out. There is an important difference between tools that block ads and pop-ups (i.e. tools that modify our perceptual environment) and tools that block particular websites from being visitable (i.e. tools that modify our field of action possibilities). An app like Self-Control blocks access to websites for a preset amount of time, even when the user tries to visit them. While the former are tools to keep *external* triggers in check, the latter are tools to restrain our *internal* triggers.

The fundamental issue seems to me to be the following: Simon (and consequently Williams) operates with a view in which the need for attention arises out of a structural bottleneck in human-information processing. This is a very common view in classical cognitive science. Attention-pioneer Broadbent writes for example: “selection takes place in order to protect a mechanism of limited capacity” ([Broadbent 1971](#), p. 178). In a 1983 chapter on attention, Marsel Mesulam takes this thought to its logical conclusion: “If the brain had infinite capacity for information processing, there would be little need for attentional mechanisms” ([Mesulam 1983](#), p. 125).³ Simon’s conception of attention is premised on a cognitive architecture in which perceptual and motor processing might be parallel, but in which central processing (i.e. cognizing or thinking) is the mechanism of limited capacity: “[h]uman beings, like contemporary computers, are essentially serial devices. They can attend to only one thing at a time” ([Simon 1971](#), p. 41). The scarce cognitive resource for Simon is the serial activity of central processing. As a result, Simon defines attention as having central processing involved in processing information.

There seem to me two obvious downsides to the attention-scarcity view. First of all, the way Williams conceives of the problem of attention-scarcity, as losing control as a result of too much information in the environment, suggests that controlling one’s attention is relatively easy in the absence of information-overload (just like controlling the placement of Tetris tiles is easy in the beginning of the game). If you moreover want to maintain that attention-scarcity is a relatively recent development, then one needs to claim that before the advent of digital technologies (or some other relevant technological development), having control over one’s own attention was relatively straightforward. This seems an implausible claim. The emphasis on meditative and attentional practices across many contemplative traditions suggests that control over one’s attention always has been a rare achievement (see [Kreiner \(2023\)](#) for an account of Christian monastic traditions in the Early Middle Ages). Leading a life of the mind (i.e. having “agentive control” over one’s attention) has never been an easy feat, regardless of the amount of

² For example, Pedersen et al. write: “Indeed, it was the psychologist Herbert Simon who first defined the basic terms of what is called the “attention economy”.” ([Pedersen et al. 2021](#), p. 311)”

³ To be fair to Mesulam, in a revised version of the book, published in 2000, the passage has disappeared and the introduction has been rewritten to reflect a much more nuanced account of the different bottlenecks involved in human cognition (cf. [Mesulam 2000](#)).

information around. It seems perhaps more plausible than that the omnipresence of digital technologies, rather than introducing distraction, made possible different and perhaps more pervasive forms of letting one's mental dynamics wander (Bruineberg and Fabry 2022).

Second, on Simon's view attention is primarily a perceptual phenomenon: there is too much information in the environment to be perceived and processed, and the challenge is to select the relevant information for further processing. If anything, the bottleneck widens again from central processing to action. It is therefore difficult to see how Simon could account for internal triggers: urges for action that arise in the absence of perceptual stimuli and that are often directed at bringing about "more" information, rather than less. It seems that we require an approach that is oriented around action and that takes as a starting point that cognitive control has always been an achievement. Fortunately, such an approach to attention is available.

Selection for action

The idea that limited capacity is explanatorily prior to selective attention is challenged by *selection for action* views of attention (Allport 1993; Neumann 1987, 1990; Wu 2011, 2014). On this alternative view, the need for attention arises because animals have at any point in time more action possibilities at their disposal than they are able to carry out. As Odmarr Neumann writes:

The problem is how to avoid the behavioral chaos that would result from an attempt to simultaneously perform all possible actions for which sufficient causes exist, i.e., that are in agreement with current motives, for which the required skills are available and that conform to the actual stimulus situation. (Neumann 1987, p. 374)

In order to avoid the "behavioral chaos" resulting from trying to do everything at once, the animal's processing needs to be selective from the start. Limiting processing in order to maintain behavioral coherence is an achievement.

The selection for action account of attention reverses the explanatory order between limited capacity and attentional selection. Odmarr Neumann and Alan Allport (Allport 1987, 1993; Neumann 1987, 1990) put their views of attention in explicit contrast with a perceptual bottleneck view of attention. On their view, attention is not a way to protect a mechanism of limited capacity, but limiting processing is a way to ensure the coherence of action selection.

The implication of Neumann's and Allport's explanatory reversal is that attention is the outcome of cognitive activity, not the cause of altered cognitive activity. Fernandez-Duque and Johnson (2002) draw a broad distinction between "Cause Theories of Attention" and "Effect Theories of Attention" (James 1890; Johnston and Dark 1986; Fernandez-Duque and Johnson 2002), each rooted in different conceptual metaphors. Cause theories hold that paying attention to a stimulus "causes" it to be processed differently from an unattended stimulus. The main question for cause theories is to give an account of what (or who) allocates attention to the relevant stimulus, while avoiding a homunculus-like regress or taking out a "loan of intelligence" (Dennett 1978; Braem and Hommel 2019). On a cause theory, attention needs to be something like "a causally efficacious substance-like reality that modulates cognitive processes" (Fernandez-Duque and Johnson 2002, p. 158). Simon's notion of attention seems to be a prime example of a cause theory of attention.

On an effect account (most prominently represented by selection-for-action theories and competition models (Desimone and Duncan 1995; Cisek 2007; Cisek and Kalaska 2010)), "what we call 'attention' is an emergent property or epiphenomenon" (Fernandez-Duque and Johnson 2002, p. 158), a byproduct of some more fundamental process. As a consequence, there is no single "attention center" in the brain from which attention originates. Attention is instead distributed throughout the cognitive system.

This little detour through debates in the psychology of attention serves to motivate the approach taken in the rest of this paper. In one of his latest writings on the topic, Alan Allport (2011) speculates that what might be seen as the "executive control center" of attention is in fact a cognitive system's increased sensitivity to context. Interestingly, Allport mentions predictive coding as a close ally of the approach to attention he advocates:

A crucial role in these [predictive coding] models is played by cortico-cortical "backward connections", as defined by their cortical layers of origin and termination. Backward connections are both more numerous and more widely branching than forward connections, and they transcend more levels, consistent with their postulated role in mediating context effects (e.g., Zeki and Shipp 1988). Moreover, context effects appear to be essentially *universal* to cortical processing: cortical units have dynamic receptive fields that can be modulated from moment to moment by activity in other units, including anatomically remote areas (McIntosh 2000; Friston 2002). Within this framework, attention becomes simply "an emergent property of 'prediction'" (Friston 2009, p. 300). (Allport 2011, p. 41)

Following Allport's analysis, predictive coding provides an "effect account" of attention in which attention is a byproduct⁴ of prediction. If attention is "an emergent property of 'prediction'" (Friston 2009, p. 300), and prediction itself is a function of both context-sensitive models of the environment and the structure of the environment, then, one should be able to analyze how the altered structure of the environment leads to different ways of predicting, and hence of different ways of attending. More than "there just being too much information," we are in a position to give a more precise and nuanced account of how attention is mediated by the omnipresence of digital technologies. In particular, I think the analysis can shed light on the agent's active involvement in bringing about particular states that, in turn, direct attention.

To manage expectations, I do not think that current active inference models can be used to model behaviors as complex as humans situated in a highly structured technological niche. What follows can hopefully be used as thinking tools and intuition pumps to better understand how human attention is structured in the attention economy.

Active inference

Active inference is a computational framework based on the premise that living systems selectively anticipate exchanges with their environment, and, in acting to realize their anticipations, achieve adaptive interactions with their environment (Clark 2015; Friston 2011; Hohwy 2013; Parr et al. 2022; Pezzulo et al. 2015).

The backbone of active inference is a synthesis of two influential ideas in the history of cognitive science: the Helmholtzian idea that error-correction serves to infer a hypothesis that best

⁴ Both Allport and Friston seem to use the term "emergent" in a slightly non-standard sense that is more akin to attention being the "consequence," or "effect," of prediction.

explains the available data (Gregory 1980; Friston et al. 2012) and the cybernetic idea that error correction serves to maintain favorable conditions (Ashby 1956; Seth 2014). According to the original Helmholtzian proposal, perception functions analogously to scientific hypothesis-testing: the agent holds a model that maps hypotheses about the world to their sensory effects. By minimizing the discrepancy (error) between predicted sensory effects and the actual sensory effects, the system will arrive at the hypothesis that best explains the data. Predictive success means epistemic success.

Cybernetics provides a control-theoretic perspective on error-correction. An agent's prolonged existence is dependent on the control of essential variables (such as body temperature and metabolic needs). When a deviation occurs (say, low blood-sugar level), the body compensates to reduce the deviation. In some cases, this can happen by internal changes (by releasing glycogen stores in the liver) or by external changes (seeking out sugar). The right kind of compensation requires a grasp on the relationship between sensed deviations and the relevant actions.

Although active inference models are sometimes introduced as a first principles theory of everything cognitive, they first emerged in cognitive science in the context of active vision: neural models underlying an agent's capacity to actively probe and make sense of a continuously changing environment (Rao and Ballard 1995; Friston et al. 2010). On this approach, active vision requires a generative model of the structure of the environment. Using the generative model, an agent can infer the "hidden" structure of the environment based on its observations. An important aspect of active vision is the selection of targets for saccadic searches. While perceiving a scene, saccading to something the agent has already seen will not deliver new information, while redirecting saccades at random will, in most instances, be sub-optimal. To efficiently deal with a changing environment, an agent needs to select saccades (or actions more generally) that resolve the greatest uncertainty about the state of the environment. For example, hearing a loud sound might lead to high uncertainty about the cause of the sound, and, at the same time, a visual orientating action toward the (estimated) source of the sound to disambiguate between different causes. Perception is therefore described in this literature as generating the most likely hypothesis that fits observations, while actions serve as "experiments" that can disambiguate between competing hypotheses (Gregory 1980; Friston et al. 2012).

The question is now how value gets incorporated in models of active vision. It is of no use to orient toward the headlights of an incoming truck, and correctly infer that a truck is about to run you over, only to be rooted to the spot. Early active inference models assumed that the agent's generative model was biased toward adaptive states (cf. Bruineberg et al. 2018a). Only if the agent's expectations of its environment are congruent with its own existence will the agent persist in reality. This "optimistic" generative model is simply assumed, motivated by evolutionary considerations (an agent that does not predict itself to stay alive will not survive for long), or learned in specific settings. For example, Friston et al. (2009) make their agent learn in a controlled environment in which the agent is shown the correct trajectory without being able to intervene. After having learned the succession of states that lead to the adaptive outcome, the agent is then put in an uncontrolled environment where it is able to act on its environment in order to make its observations congruent with the succession of states that it has previously encountered (i.e. the ones that lead to adaptive outcomes). Although this might be a good model for some forms of social learning, it fails as a general account of learning.

In contemporary models of active inference (Friston et al. 2015), value and epistemics are integrated a bit more neatly by distinguishing between the quantities of "free energy" and "expected free energy." Free energy pertains to the epistemic dimensions of active inference: its minimization involves finding the simplest hypothesis that best explains the available observations. Expected free energy has both a hedonic (i.e. value-related) and an epistemic component. In what follows, I will provide a relatively abstract and non-mathematical version of the argumentation (some of this literature tends to be a bit technical both in terms of mathematical formalisms and the concepts involved, see Friston et al. 2015; Tschantz et al., 2020; Parr et al. 2022 for details).

Expected free energy is tied to a specific policy, a course of action that the agent can undertake. It says: "what is the free-energy I expect to receive if I were to pursue this policy". Verbally, expected free energy can be written out as:

$$\text{Expected free energy} = \text{expected cost} + \text{expected ambiguity}$$

$$\text{Probability of policy } \alpha_1 / \text{expected-free energy of policy}$$

The expected cost term states: "how close will following a particular policy bring me to a desired observation." Desired observations are the way to bring (and make explicit) value and reward. If the agent's desired observation is "tasting coffee," then a policy that involves pouring oneself a cup of coffee will involve less expected cost than a policy that does not. The expected ambiguity term roughly states: "how much uncertainty will be reduced by pursuing this policy." A policy that explores a location the agent thinks it knows about will have less expected ambiguity than a location the agent is uncertain about. The probability of pursuing a policy is then inversely proportional to the relative expected free-energy of the policies: the lower the expected free energy of a policy, the higher the probability of that policy being selected. In other words: "The agent will pursue those policies most often that it expects will minimize free-energy." This involves selecting policies that are expected to lead to desired observations and selecting policies where new information is expected to be learned.

The minimization of expected free energy provides a relatively simplified, yet elegant, account of action selection in an environment that contains both uncertainty and elements that are valuable for the agent. In the following, I will go into more detail in specific aspects of the account that I think are relevant for the discussion on attention and technology that I started this paper off with.

Adversarial inference

Most of the simulations in the active inference literature assume a rather passive environment: the environment is static except for the changes an agent makes to it or it has a rather simple intrinsic dynamic (Friston et al. 2016). In some other cases, simulations involve two mutually inferring active inference agents, equipped with the same generative model, resulting in synchrony (Friston and Frith 2015). In cases where the environment changes as a function of the behavior of the agent, focus lies on beneficial forms of niche-construction, such as the formation of so-called desire paths (Bruineberg et al. 2018b).

However, it does not take much effort to think of an active environment, which itself engages in a form of active inference (see Fig. 1). One can think of the algorithms, mentioned in the introduction, as active inference agents that gather data about users and generate a user's newsfeed in order to keep them scrolling as long as possible. One can think of these algorithms as "seeing" the user's actions over time (its clicks, timing, mouse movements, perhaps even its saccades), as well as the actions

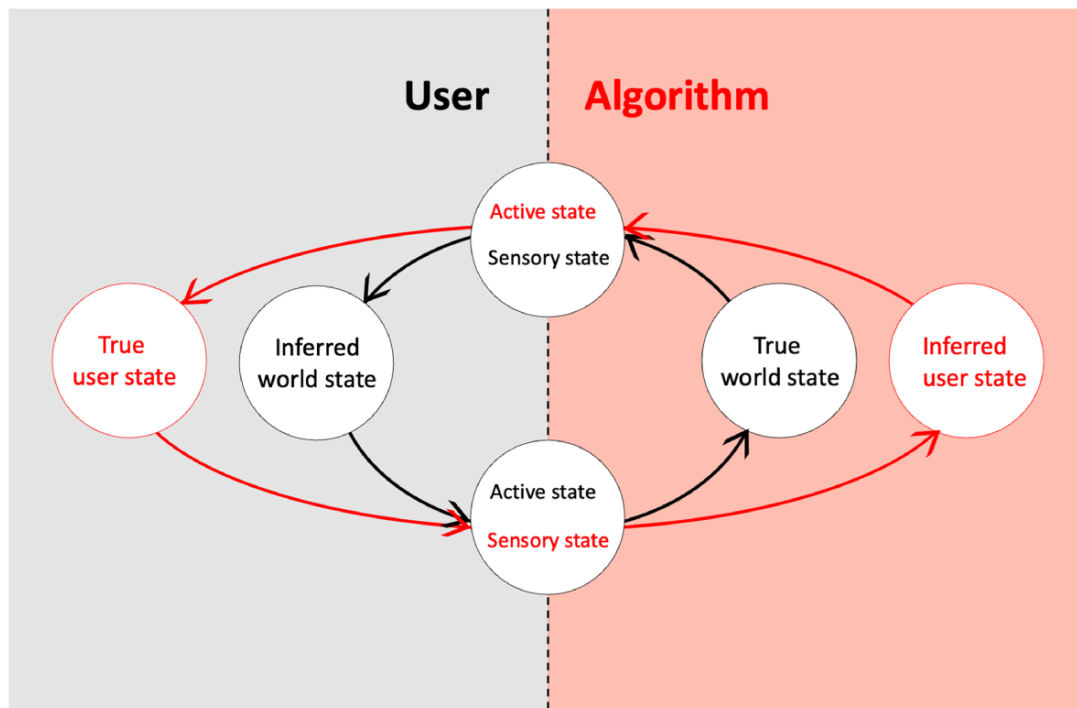


Figure 1. Schematic depiction of “adversarial inference.” The user predicts the environment, and the environment actively predicts the actions of the user. The user’s active states are the environment’s sensory states and the user’s sensory states are the environment’s active states

of a whole range of relevantly similar agents in a whole range of similar environments. The algorithm “acts” by generating the newsfeed on the user’s screen. The “desired observations” (or engagement goals) of the algorithm are, for example, to keep the agent scrolling through the timeline for as long as possible. The algorithm is doing well if it sees observations corresponding to its “goals,” and otherwise changes the world to bring it closer to these “desired” observations. If the goals of the user and algorithms are opposed (something that seems baked into the very premise of the attention economy), this qualifies as a case of adversarial inference.

The means by which an algorithm learns to keep an agent engaged can have different kinds of side-effects. For example, in an attempt to keep users watching, recommender systems can provide more and more radical content, as such contents are likely to keep the user watching (Alfano et al. 2021). Moreover, if negative mood is a predictor of prolonged social media use, downregulating the user’s mood might be an effective way to obtain desired observations (Lanier 2018).

Such fine-tuning of the reward structure in order to optimize engagement has been perfected in the gambling industry. In *Addiction by Design*, Natasha Dow Schüll (2012) describes in detail how the gambling industry developed algorithms that would gather data about a specific user in a specific session and calculate “how much that player can lose and still feel satisfied, thereby establishing personalized “pain points.” When the software senses that a player is approaching the threshold of her pain point, it dispatches a live “Luck Ambassador” to dispense rewards such as meal coupons, tickets to shows, or gambling vouchers” (Schüll 2012, p. 154).

Timms and Spurrett (2022) introduce the notion of “hostile scaffolding” to characterize environmental structures that change the cognitive demands of a task in order to serve the interests of one or more other agents. If the environmental structure is changed on the spot as a function of the agent’s behavior, then this qualifies

as what Timms and Spurrett call “deep hostile scaffolding.” If Williams’ analysis of the attention economy as a brute optimization of “engagement goals” irrespective of the consequences is correct, then the algorithms that generate a timeline based on a detailed user profile do count as an exemplary instance of deep hostile scaffolding.

All of this paints a bleak, but also quite generic, adversarial picture with respect to human–technology relationships. It should be noted that not all of the environments (digital or not) are adversarial in nature. My aim in the next section is to focus on how attention and salience are operationalized in active inference.⁵

Attention, salience, and novelty

Attention is one of the most widely used constructs in the psychological, cognitive, and neuroscientific literature. As mentioned above, one of the distinct advantages of using active inference is the ability to provide a precise mathematical operationalization of some core aspects of attention. In order to operationalize attention within active inference, more detail is needed.

So far, I have only presented prediction based on a generative model, but not the way in which the generative model is changed and updated in light of new exchanges with the environment. Key here is that the agent does not just predict specific causes of sensory observations, but a distribution of causes. For simple (Gaussian) distributions, these distributions can be captured by their “mean value” and their “variance,” or the inverse of variance, “precision.” Visually, a probability distribution which is sharply peaked around some value has low variance and high precision, while a probability distribution which is very broad

⁵ The more ambitious claims about attention made by proponents of active inference have been criticized by Ransom et al. (2017, 2020). My claim is just that aspects of attention that are relevant for the attention economy can be usefully conceptualized and modeled using active inference, not that active inference explains “everything” about attention.

has high variance and low precision. “Precision” and “variance” are both measures of the “uncertainty” related to a probability distribution.

It is possible to give something like an “anatomy of attention” based on the functional role that particular forms of uncertainty play within the active inference framework. Parr and Friston (2017) distinguish between “attention as gain” and “attention as salience”. Gain is primarily associated with regulating the uncertainty related to sensory signals. For example, in a foggy room, an agent should expect its auditory signals to be more precise than its visual signals. In estimating the state of the environment, auditory signals are inferred to be more precise than visual signals. The relative uncertainty of sensory signals plays a decisive role in weighing different forms of sensory evidence. In the foggy room, incoming visual signals have a comparatively small effect on the agent’s estimate, while auditory signals have a relatively large effect. The same mechanism of uncertainty accounts for relevant and irrelevant aspects of a stimulus. If the agent expects the letter A to appear on a screen, its expectations will involve a precise mapping between the hidden cause “letter A” and a particular visual form and shape. If the letter can take on any color, there will be a highly uncertain mapping from the hidden cause to the observed wavelength, such that observing a particular wavelength does not change the agent’s expectation of a particular letter (cf. Parr and Friston 2019).

The second notion of attention, “attention as salience” captures a more action-oriented aspect of active inference. “Attention as salience” is a property of policies. Those policies are salient that are expected to minimize free energy. Remember that expected free energy is a function of both “expected cost” and “expected ambiguity.” Those policies are salient that will bring the agent close to desired observations *and* that are expected to reduce uncertainty. In the absence of desired observations, salience is associated exclusively with selecting policies that are expected to deliver unambiguous information. In the absence of uncertainty about the environment, policy selection reduces to reward-based decision-making: selecting the policy that best delivers desired observations.

A further distinction can be made between the reduction of “certain ambiguity” and “uncertain ambiguity” (Schwartenbeck et al. 2019). Certain ambiguity refers to the reduction of ambiguity in places where it is expected to be found (i.e. hidden state exploration); uncertain ambiguity refers to the reduction of uncertainty about the statistics of the environment (i.e. model parameter exploration). For example, one can throw a die in order to figure out whether the die is fair or not (i.e. explore the model parameters), but even when one has established that the die is fair, each throw is expected to resolve ambiguity (i.e. explore the hidden state). Schwartenbeck et al. (2019) call the resolution of ambiguity about states “salience” and the resolution of ambiguity about parameters “novelty”: “‘salience is to inference’ as ‘novelty is to learning’” (p.3).

Although the distinctions introduced by Schwartenbeck et al. (2019) are useful for illuminating different forms of exploratory behavior, it creates a slight terminological tension with other discussions of salience that focus on the phenomenologically salient (or noticeable) character of particular states or actions (Archer 2022). If active inference is right, a policy’s salience in this richer sense is explained by the extent to which a policy is expected to deliver preferred observations, “salience” and “novelty”. In what follows, I will use the term salience to refer to a property of policies in this richer sense.

Following the logic of active inference, one way to attract “attention as salience” is to crank up the amount of uncertainty

in particular areas of the agent’s environment. Indeed, it seems to be the case that digital environments provide essentially limitless sources of reducible uncertainty. This is true for social media platforms that allow for infinite scrolling: content is added just as fast as the agent can scroll through its newsfeed. But also more mundane websites and applications, such as an e-mail inbox, provide highly variable *potential* novelty. At each moment in time, an important e-mail *could* be waiting in my inbox, or a highly relevant news article *could* just have dropped on the homepage of my favorite newspaper. I might just go and briefly check if I missed anything relevant. The uncertainty here seems to be mainly what Schwartenbeck et al. (2019) call “certain uncertainty.” The uncertainty does not derive from a lack of knowledge about the parameters that specify the environment but is an intrinsic and structural feature of digital environments. Important here is that the uncertainty is irreducible: even right after checking my inbox (and hence reducing uncertainty about its state), I might go straight back to check again. Aranda and Baig (2018, p. 19:3) quote from an interview they conducted among excessive smartphone users:

“When I’m bored, I keep going into my news app and tapping the same article over and over, hoping for a new story to read.”

From the perspective of active inference, checking habits are a logical consequence of living in an environment that contains information channels that provide large irreducible amounts of uncertainty.

As mentioned above, the current analysis serves as a thinking tool rather than as a concrete model for human–technology interaction. The Markov Decision Processes currently in use work well in simple environments with simple tasks, but do not scale well up to more complex behaviors. At least two questions stand in the way of scaling-up. One question pertains to policies and motor control, and the other to the nature of desired observations.

Policy selection and effort

A first question to ask is: which policies figure in the agent’s action selection process? In principle, the agent has a large (if not infinite) amount of policies available. One pragmatic option, prevalent in toy models of active inference, is to equip an agent with a limited set of policies that are relevant for its current task. Especially in MDP models, both the environments and the policies in play are rather abstract, glossing over motoric details (i.e. “move left,” “move up,” “stay put,” etc.). Although these simplifications work well enough for proof-of-principle simulations, they fall short when wanting to model ecological decision-making (Cisek and Kalaska 2010; Cisek and Pastor-Bernier 2014; Pezzulo and Cisek 2016).

In more ecological settings, the motoric details of the decision process matter. First of all, there is good evidence that action selection is a distributed process in which multiple selection processes occur at the same time (Cisek 2012). Cisek and Kalaska (2010) differentiate between “action selection,” the process of choosing between possible alternatives, and “action specification,” the process of specifying the parameters of possible action. Policy selection is hierarchical and parallel. More importantly, action specification does not just facilitate action selection (once the agent decides to reach, the parameters of its reaching are already specified), but also feeds back in action selection itself (if the reaching behavior is motorically more complex than the pointing behavior, the agent is more likely to select the pointing behavior).

The role of (expected) motoric effort in action selection has received relatively little attention in the literature on active inference.⁶ Only recently have models of decision-making started to incorporate the way (expected) effort modulates decision-making (Rangel and Hare 2010; Lepora et al. 2015; Shadmehr et al. 2016; Pierrieau et al. 2021). Pierrieau et al. (2021) hypothesize based on their findings that “motor costs quickly modulate the early formation of action representations and/or the competition process taking place in parieto-frontal regions, and thus automatically bias action selection, even when it is supposed to rely on abstract or cognitive rules” (p.2). This suggests that policy selection can be modulated by expected action costs. It seems possible to incorporate motor costs in the expected costs part of expected free energy, but, as far as I am aware, such models have not been developed.

What this points to, however, is a different lens through which the attention-grabbiness (in the salience sense) of digital technologies can be understood: digital technologies are “frictionless,” portable multi-tasking devices. They are frictionless in the sense that their use typically requires very simple and smooth motor actions (such as tapping and swiping), typically with one or more fingers. Because they are portable, they offer their action possibilities everywhere and anytime regardless of place, time, or context. Moreover, because they are multi-tasking devices, they can be used for many different activities. Switching between different activities often requires just as much effort as continuing to do one and the same activity. For example, scrolling through the text I am writing and inserting a word requires a two-fingered swipe and a small number of taps on the keyboard. Switching from writing to scrolling through Twitter requires a four-fingered swipe and a tap on the mousepad followed by a couple of taps on the keyboard. Looking up a chapter I cite in my reference manager is motorically just as complex as starting a completely unrelated task altogether.

The above suggests that if action selection is modulated by expected effort, and if switching to distracting activities is motorically easier (or at least comparable in complexity) to staying on task, then the majority of our digital devices are set up to bias action selection toward distraction (or at least not away from it). This is not a case of hostile scaffolding in Timms and Spurrett’s (2022) sense of above, but just “bad scaffolding.” I will return to this point in the conclusion,

Motivation and control

The last point I want to consider is the nature of goals and desired observations in active inference. So far, I have simply assumed that an agent has “desired observations” that serve as a standard to measure the expected cost of particular policies (i.e. which of these policies brings me closest to my desired observations). In simple decision-making tasks taking desired observations for granted might be unproblematic: we might just assume that a particular observation (seeing the green dot at the end of the maze) is rewarding. In ecological settings, however, an agent typically has multiple “desires” at the same time at multiple levels of abstraction, where the relevance of each of those desires is itself dependent on current observations. This gives rise to what Pezzulo et al. (2018) call a “multidimensional drive-to-goal problem.” For example, if I am (following the logic of expected free-energy minimization) on my way to the bakery to satisfy my desired observation of tasting pastry, I might happen to come across an old friend. In such a case, I should be able to revise my

desired observation and prioritize my (until then latent) desire to see my friend, such that I start talking to her, instead of passing her by and racing to the bakery.⁷ The sudden achievability of a different desire than the one that is currently driving my behavior should make me switch my priorities. Hence, goals and desires are themselves adaptive, and their relevance is dependent on context.

From the perspective of active inference, the relevance of a particular desired observation is given by its associated precision (Pezzulo et al. 2018). A precise expectation of tasting pastry will guide policy selection toward a bakery, sidestepping any possibility to resolve ambiguity along the way. Pezzulo et al. (2018) propose to model multidimensional drive-to-goal problems by means of a theory of “motivated control.” The authors propose two functionally separate but interacting hierarchies: a control hierarchy and a motivation hierarchy. The control hierarchy captures the causal regularities in the environment ranging in complexity from sensorimotor contingencies to more abstract beliefs like knowing there is typically a line at the bakery on Saturday mornings. The motivation hierarchy ranges in complexity from a craving for carbs, to wanting to eat less sugar, to more abstract desires like wanting to be a sociable person.

At the basic (sensorimotor) level, action policies compete with one another, and their competition is biased (or contextualized) by higher level states (themselves in competition with one another). The resulting choice architecture has strong affinities with earlier mentioned competition models (Desimone and Duncan 1995; Cisek 2007; Cisek and Kalaska 2010).

The novelty of Pezzulo et al. (2018) with respect to these earlier models is the way the authors make explicit the functional integration between the control hierarchy and the motivation hierarchy. Motivational states like “I should eat less sugar” will, if inferred to be relevant, bias expectations away from going to the bakery. Importantly, if active inference is right, the higher-level contextual goal-states do not come from above, but are themselves learned and inferred, in part based on behavior. As a consequence, the authors propose that making progress toward a goal (i.e. approaching the bakery) increases the precision of beliefs about policies that achieve the goal (i.e. “apparently I want to get pastry”). The authors write:

Thus, when precision is itself inferred, successful goal-directed behaviour creates a form of positive feedback between control and motivational processes. [...] Intuitively, it is sometimes difficult to start a new task, but once progress has been made, it becomes difficult to give it up – even when the reward is small. A possible explanation is that, as goal proximity increases, its inferred achievability increases – with precision – hence placing a premium on the policy above and beyond of its pragmatic value. (Pezzulo et al. 2018, p. 302)

The loopy dynamics between control and motivation constitutes an intricate case of motivated reasoning, in which the agent both starts to do what it wants but also starts to want what is likely to happen: I want the pastry in part “because” it is in front of my nose.

The upshot of the motivated control model is that both “achievability” and “goal-proximity” bias salience. An agent’s action selection is primed toward what is achievable. This is a mechanism that allows an agent to effectively switch courses of action: the

⁶ A first step has been recently made by Parr et al. (2023), who use active inference to model “cognitive” effort, operationalized as the effort to resist a mental habit.

⁷ When taken in at full force, the problematic amounts to the “frame problem” or the “relevance problem” (Dreyfus 1992; Wheeler 2008; Rietveld 2012; 2016; Shanahan 2016).

immediate achievability of talking to my friend increases the motivational precision of wanting to be a sociable person, and decreases the precision of craving sugar, constituting an attentional shift in the relevance of particular motivational states. This capacity to interrupt what one is currently doing in order to satisfy a more urgent or achievable goal seems central to the motivational economy of any agent that needs to satisfy multiple goals or needs over time (Simon 1994). However, the same mechanism allows for distraction: a biasing away from what one is currently doing, and toward an easier achievable activity with its own kind of “reward.”

Remember that in Eyal’s “Hooked” model (Eyal 2014), the ultimate aim is to create an association between a user’s need and the use of a company’s product as a source of relief. In the context of active inference, this amounts to a motivational state that, when inferred, directly biases action selection toward particular policies. An association between, say, anxiety about social status and Twitter-likes, will, whenever the anxiety plays up, lead to precise expectations about the policies that bring about Twitter-scrolling. More speculatively (but in line with the motivated control model), prolonged Twitter use will in turn enhance the relevance of the motivational state (“If I spent so much time on Twitter, apparently social status is important to me”). At any rate, the learned associations between motivational states and action policies forms what Eyal calls an “internal trigger”: an urge to act that is not directly prompted by any aspect of the environment.

Conclusion

In his book *Stand Out of Our Light*, James Williams (2018) calls the liberation of human attention the “defining moral and political struggle of our time” (xii). Attention is intimately tied to notions such as agency, autonomy, and self-determination. I agree. It is therefore of utmost importance that the debate on the attention economy operates with a plausible account of attention, and with a specific understanding of what it is about our current environments that undermines attention.

In presenting the current debate on the attention economy, I have articulated three desiderata of a theory of attention (wrapped up as criticisms of the dominant account). First of all, any theory of attention should avoid “attention nostalgia,” the idea that in the information-scarce days (before mass media) controlling one’s attention was easy. This was evidently not the case (Kreiner 2023). Second, a theory of attention should be able to articulate what it is about our current environments that makes paying attention to the relevant things more difficult. The information-scarcity account points to the sheer “amount” of information in our environments that makes attention more difficult to control, due to added “pressure” on the bottleneck from perception to central cognition. And third, an account of attention should be able to accommodate the fact that many instances of distraction are not driven by salient stimuli. Models of digital technology-related habits emphasize the importance of “internal triggers,” urges to bring about an application in the absence of stimuli.

Selection for action models operate with an account according to which attention is tightly integrated with an agent’s ongoing interactions with its environment. Attention is said to be an “emergent property” or “outcome” of a more fundamental selection process. In particular, active inference provides a set of tools to develop an “anatomy” of different forms of attention. One central distinction in this literature is between “attention as gain” and “attention as salience.” While “attention as gain” involves the weighting of evidence in updating one’s current hypothesis, “attention as salience” is central to action selection, and it is here that active inference has a number of insights to offer.

First of all, active inference can make sense of explicitly adversarial or hostile scenarios in which the environment actively infers the state of the agent to provide it with content that satisfies the environment’s (and not the agent’s) goals. Such cases of adversarial inference well documented in the gambling industry, and the timelines presented by digital environments provide plausible candidates for such hostile scaffolding.

But active inference provides insight into more ways in which digital environments reorganize “attention as salience.” In this paper, I have focused on three: I have focused on the central role that novelty and the expectation of novelty plays in action selection. Second, I have focused on the role that motor costs, or friction, plays in action selection: action selection is biased toward actions that involve less expected effort. The third insight pertains to the way in which the motivation and achievability of actions interact. The salience of actions is dependent on one’s current motivational state, but the motivational state is itself dependent on the perception of the environment and on one’s current activities.

Taken together, these aspects of active inference can make intelligible that, besides the overtly indubitably adversarial and hostile aspects of digital engagement, there is a profound mismatch between the kind of agent’s that we are, and the kind of environments that we inhabit. As active inference agents, we are not well “equipped” to deal with environments that present frictionless engagement (including frictionless task-switching), practically infinite amounts of easily accessible novelty, highly variable reward, and content that taps into different motivations. The challenges users encounter with digital technologies are therefore not specifically about information-overload or irresistible content, but more about the continuous availability of easily available novelty and achievable rewards.⁸

To contrast, there is a certain way of thinking about agency and action selection on which carrying low-effort, maximum-possibility devices with us all day makes a lot of sense. If I know firmly what I want to achieve, then a device that helps me realize my preferences with as little effort as possible will generally be a good thing. Jan Slaby dubs such thinking about the mind a “user/resource model”:

Baseline mentality in many of the example cases under discussion is that of a fully conscious individual cognizer (“user”) who sets about pursuing a well-defined task through intentional employment of a piece of equipment or exploitation of an environmental structure (“resource”). (Slaby 2016, p. 5)

Following this model, there is a clear division of labor between user and technology: the “user” brings the goals, and the technology simplifies the means by which the goals can be achieved. In some cases, devices are designed and (particularly) branded as “autonomy-enhancing” exactly because they minimize the effort it takes to perform a task.⁹

If active inference is right, we are not such agents.

⁸ One reviewer is right to point out that information-overload is not necessarily at odds with the current account. After all, environments with large amounts of novelty will also need to be rich in information. I hope to have made clear that the decisive factor for attention and distraction is the structure and dynamics of information, rather than the sheer amount. Libraries are filled with information, but this information is not structured in a way that fosters distraction: it does not make sense to develop a checking habit for a particular page of a book.

⁹ See Tim Wu’s (2017) analysis of the remote control and the emergence of channel surfing for an especially vivid example.

Acknowledgements

I would like to thank Alistair Isaac, Colin Klein, Odysseus Stone, the members of the Monash Centre for Consciousness & Contemplative Studies (in particular, Andrew Corcoran, Mengting Zhang, and Jakob Hohwy), as well as two anonymous reviewers for providing critical comments on earlier drafts of this paper. This research was generously supported by a Macquarie University Research Fellowship.

Conflict of interest

None declared.

Data availability

No data was used for this paper.

References

- Aagaard J. Digital akrasia: a qualitative study of phubbing. *AI Soc* 2020;**35**:237–44.
- Alfano M, Fard AE, Carter JA et al. Technologically scaffolded atypical cognition: the case of YouTube's recommender system. *Synthese* 2021;**199**:835–58.
- Allport A. Selection for action: some behavioral and neurophysiological considerations of attention and action. In: Heuer H, Sanders A (eds.), *Perspectives on Perception and Action*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1987, 395–420.
- Allport A. Attention and control: have we been asking the wrong questions? A critical review of twenty-five years. In: Meyer DE, Kornblum S (eds.), *Attention and Performance 14: Synergies in Experimental Psychology, Artificial Intelligence, and Cognitive Neuroscience*. Cambridge, MA: The MIT Press, 1993, 183–218.
- Allport A. Attention and integration. In: Mole C, Smithies D, Wu W (eds.), *Attention: Philosophical and Psychological Essays*. New York: Oxford University Press, 2011, 24–59.
- Aranda JH, Baig S. Toward JOMO: the joy of missing out and the freedom of disconnecting. *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*, Barcelona, Spain, 2018, 1–8.
- Archer S. *Saliency: A Philosophical Inquiry*. Routledge, 2022.
- Ashby WR. *An Introduction to Cybernetics*. London: Chapman & Hall, 1956.
- Braem S, Hommel B. Executive functions are cognitive gadgets. *Behav Brain Sci* 2019;**42**:e173–e173.
- Broadbent DE. *Decision and Stress*. Cambridge, MA: Academic Press, 1971.
- Bruineberg J, Fabry R. Extended mind-wandering. *Philos Mind Sci* 2022;**3**:1–30.
- Bruineberg J, Kiverstein J, Rietveld E. The anticipating brain is not a scientist: the free-energy principle from an ecological-enactive perspective. *Synthese* 2018a;**195**:2417–44.
- Bruineberg J, Rietveld E, Parr T et al. Free-energy minimization in joint agent-environment systems: a niche construction perspective. *J Theor Biol* 2018b;**455**:161–78.
- Castro C, Pham AK. Is the attention economy noxious? *Philos Impr* 2020;**20**.
- Chomanski B. Mental integrity in the attention economy: in search of the right to attention. *Neuroethics* 2023;**16**:8.
- Cisek P. Cortical mechanisms of action selection: the affordance competition hypothesis. *Philos Trans R Soc Lond B: Biol Sci* 2007;**362**:1585–99.
- Cisek P. Making decisions through a distributed consensus. *Curr Opin Neurobiol* 2012;**22**:927–36.
- Cisek P, Kalaska JF. Neural mechanisms for interacting with a world full of action choices. *Annu Rev Neurosci* 2010;**33**:269–98.
- Cisek P, Pastor-Bernier A. On the challenges and mechanisms of embodied decisions. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 2014;**369**:20130479.
- Clark A. *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford: Oxford University Press, 2015.
- Dennett DC (1978). *Brainstorms: philosophical essays on mind and psychology*.
- Desimone R, Duncan J. Neural mechanisms of selective visual attention. *Annu Rev Neurosci* 1995;**18**:193–222.
- Dreyfus HL. *What Computers Still Can't Do: A Critique of Artificial Reason*. Boston, MA: MIT press, 1992.
- Eyal N (2012). *The Hooked Model: How to Manufacture Desire in 4 Steps*. Nir and Far. <https://www.nirandfar.com/how-to-manufacture-desire/> (3 August 2023, date last accessed).
- Eyal N. *Hooked: How to Build Habit-Forming Products*. Penguin, 2014.
- Fernandez-Duque D, Johnson ML. Cause and effect theories of attention: the role of conceptual metaphors. *Rev Gen Psychol* 2002;**6**:153–65.
- Friston K. Beyond Phrenology: What can neuroimaging tell us about distributed circuitry? *Annual Review of Neuroscience* 2002;**25**:221–50.
- Friston K. The free-energy principle: a rough guide to the brain? *Trends Cogn Sci* 2009;**13**:293–301.
- Friston KJ. Embodied inference: or I think therefore I am, if I am what I think. In: Tschacher W, Bergomi C (eds.), *The implications of embodiment: Cognition and communication*, Imprint Academic, 2011, 89–125.
- Friston KJ, Adams R, Perrinet L et al. Perceptions as hypotheses: saccades as experiments. *Front Psychol* 2012;**3**:151.
- Friston KJ, Daunizeau J, Kiebel SJ. Reinforcement learning or active inference? *PLOS ONE* 2009;**4**:e6421.
- Friston KJ, Daunizeau J, Kilner J et al. Action and behavior: a free-energy formulation. *Biol Cybern* 2010;**102**:227–60.
- Friston KJ, Fitzgerald T, Rigoli F et al. Active inference and learning. *Neurosci Biobehav Rev* 2016;**68**:862–79.
- Friston KJ, Frith C. A duet for one. *Conscious Cogn* 2015;**36**:390–405.
- Friston KJ, Rigoli F, Ognibene D et al. Active inference and epistemic value. *Cogn Neurosci* 2015;**6**:187–214.
- Gendler TS. Alief in action (and reaction). *Mind Lang* 2008;**23**:552–85.
- Gregory RL. Perceptions as hypotheses. *Philos Trans R Soc Lond B: Biol Sci* 1980;**290**:181–97.
- Harris T. *Optimizing for Engagement: Understanding the Use of Persuasive Technology on Internet Platforms*. US Senate Committee on Commerce, Science, & Transportation, 2019.
- Hohwy J. *The Predictive Mind*. Oxford: Oxford University Press, 2013.
- James W. *The Principles of Psychology*. New York: Henry Holt, 1890.
- Johnston WA, Dark VJ. Selective attention. *Ann Rev Psychol* 1986;**37**:43–75.
- Kreiner J. *The Wandering Mind: What Medieval Monks Tell Us About Distraction*. New York: Liveright, 2023.
- Lanier J. *Ten Arguments for Deleting Your Social Media Accounts Right Now*. New York: Henry Holt, 2018.
- Lepora NF, Pezzulo G, Daunizeau J. Embodied choice: how action influences perceptual decision making. *PLoS Comput Biol* 2015;**11**:e1004110.
- McIntosh MR. Towards a network theory of cognition. *Neural Networks* 2000;**13**:861–70.
- Mesulam -M-M. Attention, confusional states, and neglect. In: Mesulam -M-M (ed.), *Principles of Behavioral Neurology*. Philadelphia, PA: F.A. Davis, 1983, 125–68.

- Mesulam -M-M. Attention, confusional states, and neglect. In: Mesulam -M-M (ed.), *Principles of Behavioral and Cognitive Neurology*. Oxford: Oxford University Press, 2000, 174–256.
- Neumann O. Beyond capacity: a functional view of attention. In: Heuer H, Sanders AF (eds.), *Perspectives on Perception and Action*. Hillsdale, NJ: Erlbaum, 1987, 361–93.
- Neumann O. Visual attention and action. In: Neumann O, Prinz W (eds.), *Relationships between Perception and Action*. Berlin: Springer, 1990, 227–67.
- Nguyen CT. How twitter gamifies communication. In: Lackey J (ed.), *Applied Epistemology*. Oxford: Oxford University Press, 2021.
- Orlowski J. Director *The Social Dilemma*. Documentary Exposure Labs, 2020.
- Oulasvirta A, Rattenbury T, Ma L et al. Habits make smartphone use more pervasive. *Pers Ubiquitous Comput* 2012;**16**:105–14.
- Parr T, Friston KJ. Working memory, attention, and salience in active inference. *Sci Rep* 2017;**7**:14678.
- Parr T, Friston KJ. Attention or salience? *Curr Opin Psychol* 2019;**29**:1–5.
- Parr T, Holmes E, Friston KJ et al. Cognitive effort and active inference. *Neuropsychologia* 2023;**184**:108562.
- Parr T, Pezzulo G, Friston KJ. *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior*. Cambridge, MA: MIT Press, 2022.
- Pedersen MA, Albris K, Seaver N. The political economy of attention. *Annu Rev Anthropol* 2021;**50**:309–25.
- Pezzulo G, Cisek P. Navigating the affordance landscape: feedback control as a process model of behavior and cognition. *Trends Cogn Sci* 2016;**20**:414–24.
- Pezzulo G, Rigoli F, Friston K. Active inference, homeostatic regulation and adaptive behavioural control. *Prog Neurobiol* 2015;**134**:17–35.
- Pezzulo G, Rigoli F, Friston KJ. Hierarchical active inference: a theory of motivated control. *Trends Cogn Sci* 2018;**22**:294–306.
- Pierriau E, Lepage J-F, Bernier P-M. Action costs rapidly and automatically interfere with reward-based decision-making in a reaching task. *ENeuro* 2021;**8**:ENEURO.0247–21.2021.
- Rangel A, Hare T. Neural computations associated with goal-directed choice. *Curr Opin Neurobiol* 2010;**20**:262–70.
- Ransom M, Fazelpour S, Markovic J et al. Affect-biased attention and predictive processing. *Cognition* 2020;**203**:104370.
- Ransom M, Fazelpour S, Mole C. Attention in the predictive mind. *Conscious Cogn* 2017;**47**:99–112.
- Rao RPN, Ballard DH. An active vision architecture based on iconic representations. *Artif Intell* 1995;**78**:461–505.
- Rietveld E. Context-switching and responsiveness to real relevance. In: Kiverstein J, Wheeler M (eds.), *Heidegger and Cognitive Science*. New York: Palgrave Macmillan, 2012, 105–34.
- Schüll ND. *Addiction by Design*. Princeton: Princeton University Press, 2012.
- Schwartenbeck P, Passecker J, Hauser TU et al. Computational mechanisms of curiosity and goal-directed exploration. *ELife* 2019;**8**:e41703.
- Seth AK. The cybernetic Bayesian brain: from interoceptive inference to sensorimotor contingencies Windt JM, Metzinger T *Open MIND*. Frankfurt am Main: MIND Group, 2014, 9–24.
- Shadmehr R, Huang HJ, Ahmed AA. A representation of effort in decision-making and motor control. *Curr Biol* 2016;**26**:1929–34.
- Shanahan M. The frame problem. In: Zalta EN (ed.), *The Stanford Encyclopedia of Philosophy*. CA: Metaphysics Research Lab, Stanford University, 2016.
- Simon HA. Designing organizations for an information-rich world. In: Greenberger M (ed.), *Computers, Communication, and the Public Interest*. Baltimore: Johns Hopkins University Press, 1971, 38–52.
- Simon HA. The bottleneck of attention: connecting thought with motivation. In: Spaulding WV (ed.) *Integrative Views of Motivation, Cognition, and Emotion*. Lincoln, NE: University of Nebraska Press, 1994, 1–21.
- Slaby J. Mind invasion: situated affectivity and the corporate life hack. *Front Psychol* 2016;**7**:266.
- Timms R, Spurrett D. Hostile Scaffolding. *Philosophical Papers*, 2023;**52**.
- Tschantz A, Seth AK, Buckley CL. Learning action-oriented models through active inference. *PLoS Computational Biology* 2020;**16.4**:e1007805.
- Wheeler M. Cognition in context: phenomenology, situated robotics and the frame problem. *Int J Philos Stud* 2008;**16**:323–49.
- Williams J. *Stand Out of Our Light: Freedom and Resistance in the Attention Economy*. Cambridge, MA: Cambridge University Press, 2018.
- Wu W. Confronting many-many problems: attention and agentic control. *Noûs* 2011;**45**:50–76.
- Wu W. *Attention*. London: Taylor & Francis, 2014.
- Wu T. *The Attention Merchants: The Epic Scramble to Get inside Our Heads*. Vintage, 2017.
- Zeki S, Shipp S. The functional logic of cortical connections. *Nature* 1988;**335**:311–7.
- Zuboff S. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London: Profile Books, 2019.