VILNIUS UNIVERSITY
FACULTY OF MEDICINE

Medicine

Department of Anatomy, Histology and Anthropology

Tautvydas Rauba, 6th year, Group 14

INTEGRATED STUDY MASTER'S FINAL THESIS

Nesąmoningas šališkumas evoliuciniu požiūriu. Literatūros apžvalga

Unconscious Bias from an Evolutionary Perspective. Literature Review

Supervisor *Assoc. Prof. Rūta Vosyliūtė, PhD*

Head of the department *Prof. Janina Tutkuvienė, PhD*

Vilnius, 2026

tautvydas.rauba@mf.stud.vu.lt

TABLE OF CONTENTS

1. SUMMARY

Background. Unconscious bias is a pervasive phenomenon that compromises health outcomes and perpetuates societal disparities. Despite its critical impact, the foundational mechanisms driving unconscious biases—specifically their evolutionary and cognitive origins—remain contested in contemporary literature, complicating efforts to mitigate them.

Aim. To review unconscious bias by defining its theoretical foundations, examining its impact on health and healthcare, and evaluating its possible evolutionary origins.

Objectives. 1) To establish a conceptual framework for unconscious bias, including its definition, related terminology, and a typology of its forms. 2) To detail the impact of unconscious bias on health and healthcare. 3) To critically evaluate evolutionary perspectives of unconscious bias by synthesizing findings from neuroscience, cognitive science, evolutionary psychology, and philosophy of science.

Methods. A narrative literature review was conducted using PubMed, PsycINFO, PhilPapers and Google Scholar, supplemented by the Stanford Encyclopedia of Philosophy and StatPearls. Search included combinations of keywords related to evolutionary theory, cognitive architecture, unconscious bias, and social cognition. In total 129 articles were included in the review.

Results. Unconscious biases are automatic associations that happen without conscious awareness. They influence how people see the world and how they behave, and they often lead to discrimination. Because biases are common and hard to change, some researchers think they may have evolved as part of human psychology. There are two main ways to explain the origins of unconscious biases. The first, domain-specific approaches, suggest that these biases are evolved adaptations. In the past, they may have helped people deal with recurring problems such as avoiding disease, detecting danger, or protecting resources. The second, domain-general perspective, argues that unconscious biases are by-products of basic mental processes like learning from experience, grouping things into categories, and making predictions. These processes can lead to biased thinking, especially in environments where social biases already exist.

Conclusions. This literature review examined the concept of unconscious bias and its evolutionary explanations. The findings indicate that the concept is theorized inconsistently throughout the literature. Moreover, unconscious bias remains a pervasive phenomenon that poses significant public health challenges. Finally, existing evolutionary accounts lack sufficient theoretical and empirical validity, and a domain-general perspective provides a plausible explanation for the origins of unconscious bias.

2. SANTRAUKA

Darbo aktualumas. Nesąmoningas šališkumas yra plačiai paplitęs fenomenas, darantis neigiamą įtaką visuomenės sveikatai ir didinantis socialinę nelygybę. Nepaisant jo svarbaus poveikio, pagrindiniai nesąmoningą šališkumą lemiantys mechanizmai (ypač jo evoliucinė ir kognityvinė kilmė) šiuolaikinėje literatūroje išlieka ginčytini, ir tai apsunkina sprendimų pasirinkimą jo mažinimui.

Tyrimo tikslas. Apžvelgti mokslinę literatūrą apie nesąmoningą šališkumą, apibrėžiant jo teorinius pagrindus, nagrinėjant poveikį sveikatai ir sveikatos priežiūrai bei įvertinant galimas jo evoliucines ištakas.

Uždaviniai. 1) Sukurti nesąmoningo šališkumo konceptualųjį pagrindą, įskaitant jo apibrėžimą, susijusią terminiją ir formų klasifikaciją. 2) Išsamiai aprašyti nesąmoningo šališkumo poveikį sveikatai ir sveikatos priežiūrai. 3) Kritiškai įvertinti nesąmoningo šališkumo evoliucines perspektyvas, apibendrinant neuromokslų, kognityvinių mokslų, evoliucinės psichologijos ir mokslo filosofijos tyrimus.

Tyrimo metodai. Naratyvioji literatūros apžvalga atlikta naudojant „PubMed“, „PsycINFO“, PhilPapers, ir „Google Scholar“ duomenų bazes, kurios buvo papildytos „Stanford Encyclopedia of Philosophy“ ir StatPearls šaltiniais. Paieška apėmė raktinių žodžių, susijusių su evoliucijos teorija, kognityvine architektūra, nesąmoningu šališkumu ir socialiniu pažinimu, kombinacijas. Iš viso 129 straipnsiai buvo įtraukti į apžvalgą.

Literatūros apžvalgos rezultatai. Nesąmoningas šališkumas (tai yra automatinės asociacijos, susidarančios be sąmoningo suvokimo) daro įtaką žmonių suvokimui ir veiksmams, bei dažnai nulemia diskriminacinį elgesį. Kadangi šališkumas yra plačiai paplitęs, dalis tyrėjų teigia, kad jis galėjo išsivystyti kaip evoliucinė žmogaus psichologijos dalis. Egzistuoja du pagrindiniai būdai paaiškinti nesąmoningo šališkumo kilmę. Pirmasis (specifiškos srities požiūris) teigia, kad šie šališkumai yra evoliucinės adaptacijos. Praeityje jie galėjo padėti žmonėms spręsti pasikartojančias problemas, tokias kaip ligų išvengimas, pavojaus aptikimas ar išteklių apsauga. Antrasis (bendrosios srities požiūris) argumentuoja, kad numanomas šališkumas yra bazinių psichikos procesų, tokių kaip mokymasis iš patirties, objektų grupavimas į kategorijas ir prognozių kūrimas, šalutinis produktas. Šie procesai gali lemti šališką mąstymą, ypač tose aplinkose, kuriose socialinis šališkumas jau egzistuoja.

Išvados. Šioje literatūros apžvalgoje išnagrinėta nesąmoningo šališkumo samprata ir jo evoliuciniai paaiškinimai. Rezultatai rodo, kad literatūroje ši samprata teorizuojama nenuosekliai. Be to, nesąmoningas šališkumas išlieka plačiai paplitusiu reiškiniu, keliančiu reikšmingų iššūkių visuomenės gerovei. Galiausiai, evoliuciniams paaiškinimams trūksta pakankamo teorinio ir empirinio pagrįstumo, o bendrosios srities perspektyva suteikia labiau tikėtiną nesąmoningo šališkumo kilmės paaiškinimą.

3. KEYWORDS

Implicit social cognition, unconscious bias, evolutionary psychology, domain-specific, domain-general, predictive processing.

4. INTRODUCTION

While some philosophers defended prejudice as an indispensable human faculty - though in a sense distinct from its pejorative meaning (see Box 1) - many Enlightenment thinkers emphatically embraced objective reason as the ultimate arbiter of truth [1]. Rationality was put forth as the standard for human conduct, and prejudice was cast away as a remnant of the past. These rationalist and egalitarian notions, among other factors, have taken root to the extent that stereotype-derived judgments are now broadly discouraged [2]. In turn, many sincerely believe they have rid themselves of all traces of biased reasoning, and the notable decrease in explicitly expressed prejudice [3] certainly appears to attest to that. However, this picture of an overall reduction in prejudice is challenged by the high prevalence of implicit bias [4]. This discrepancy is largely attributed to the fact that explicit measures of bias are compromised by social desirability effects, as individuals suppress prejudiced responses to adhere to societal norms [5,6]. In light of this, most research [7] has pivoted toward methodologies less prone to subjective distortion. Through the use of indirect cognitive and neuroimaging tools [8], *unconscious* bias became the focus of contemporary scientific studies.

Operating mostly beyond the awareness and direct control of individuals, unconscious bias is a vexing process. It exerts a wide array of negative effects on society [9], ranging from distressing interpersonal interactions and constrained employment opportunities to grave injustices in the legal and law enforcement systems and fatal errors in the health sector. While extensive effort is expended to curtail these harmful influences, unconscious bias remains highly resistant to change, with most standard interventions yielding negligible long-term effects [10]. Such pervasiveness suggests robust evolutionary underpinnings. This review investigates these claims by outlining several evolutionary accounts and evaluating their validity. Before doing so, it first thoroughly explores the concept of unconscious bias and details its impact on clinical and health outcomes.

Box 1. Two philosophers defend prejudice

Writing during the early days of the French Revolution in 1790, political philosopher Edmund Burke, the "father of modern conservatism," actively defended prejudice as a valid and valuable preconceived opinion [11]. He believed it accurately embodies a country's customs and norms, offering a reliable method for making consistent judgments in line with one's virtues. Prejudice also proves useful under pressing time constraints (i.e., "emergencies") when decisions need to be made quickly. Predictably, he dismissed "naked reason" as an insufficient guide for human conduct.

In his magnum opus "Truth and Method", written in 1960, philosopher Hans-Georg Gadamer approached the concept of prejudice by invoking the older, positive conception of the term: literally, a "pre-judgment" [12]. He claimed that prejudices constitute a "fore-structure" for understanding. In other words, rather than impairing and obstructing comprehension, prejudgments" are anticipatory structures that make meaning-making possible at all. For anything to be understood, it must represent a complete, coherent whole. Understanding requires 'anticipation of completeness', and prejudices fulfil this function by providing preliminary sketches. Gadamer conceded that while prejudices enable comprehension, they occasionally distort reality. By identifying pre-judgements that contradict other forms of knowledge, initial preliminary impressions can be updated, allowing for a more nuanced understanding to emerge [12].

Burke's characterization of prejudice closely resembles the modern conception of heuristics and their utility, although contemporary discourse discourages such reliance on them. Gadamer's approach, however, is much more nuanced. It provides a structural-level explanation that resonates with current cognitive science models, such as predictive processing [13], which suggests that the brain generates top-down predictions to make sense of incoming sensory data, or research in social neuroscience [14], which shows how split-second visual perceptions are influenced by higher-order constructs (e.g., stereotypes and prejudices).

5. AIM AND OBJECTIVES

Aim. To review unconscious bias by defining its theoretical foundations, examining its impact on health and healthcare, and evaluating its possible evolutionary origins.

Objectives:

1) To establish a conceptual framework for unconscious bias, including its definition, related terminology, and a typology of its forms.
2) To detail the impact of unconscious bias on health and healthcare.
3) To critically evaluate evolutionary perspectives of unconscious bias by synthesizing findings from neuroscience, cognitive science, evolutionary psychology, and philosophy of science.

6. METHODS

This thesis is based on a narrative literature review. The narrative approach was chosen due to the highly diverse literature on the topic, spanning multiple scientific fields, including cognitive science, neuroscience, psychology, and evolutionary psychology. This method enabled a comprehensive analysis of competing theories to produce an accurate overview of the concept of unconscious bias and its foundations.

Literature was gathered using academic databases and search engines, primarily PubMed, PsycINFO, Google Scholar, PhilPapers and Google. The Stanford Encyclopedia of Philosophy provided broad overview articles on topics such as implicit bias and evolutionary psychology. StatPearls was used for review articles on the impact of implicit biases on healthcare. The initial search relied on keyword combinations such as "implicit bias overview," "neuroscience of prejudice," "evolutionary approaches to stereotyping," and "predictive processing bias". Foundational articles were identified through the initial keyword search. Complementary studies were found by reviewing the reference lists of the main articles. Additional literature was identified through targeted searches as specific subtopics came up. Articles were selected based on direct relevance rather than strict publication dates, resulting in a final inclusion of 129 studies. Citations were managed using Zotero.

The research was organized into four main phases. The first phase consisted of gathering literature to define unconscious bias and its psychological components. The second phase focused on providing an overview of its impact on health and healthcare systems. The third phase involved identifying literature on evolutionary explanations for the development of biases. Finally, the fourth phase consisted of synthesizing critical reviews of these evolutionary theories and exploring predictive processing as an alternative explanation.

Generative artificial intelligence (Gemini 3.1 Pro) was utilized strictly for grammatical error correction and readability suggestions.

7. RESULTS AND DISCUSSION

7.1. The concept of unconscious bias

7.1.1. Theoretical background

Folk psychology refers to intuitive assumptions non-experts hold regarding social behavior and mental states [15], such as the belief that human behavior is primarily driven by explicit beliefs and conscious intentions. In contrast, the scientific perspective offers a markedly different account. Consciousness, or short-term 'working memory', is an excessively costly cognitive resource. As a result, there is substantial pressure to encode and store relevant information into unconscious, or non-declarative, long-term memory to automate behavior [16]. It is estimated that only 5% of goal-directed action is consciously controlled [17]. Conversely, approximately 95% of all mental activity consists of unconscious processing [18]. This is the context within which the primary subject of this literature review, unconscious bias, is situated.

7.1.2. Conceptualizing unconscious bias

Unconscious biases are automatic associations that occur without conscious awareness. These biases influence how people perceive the world and how they behave, potentially leading to discriminatory outcomes. They stand in contrast to explicit biases, which refer to consciously endorsed beliefs and attitudes that might manifest as deliberate discrimination. Thus, the key relevance of unconscious bias lies in its unintentional character, which nevertheless causes discriminatory behavior that the perpetrator is unaware of committing. In the cognitive sciences and social psychology, the standard technical term frequently used interchangeably with this phenomenon is "implicit bias". Because much of the existing literature treats these terms as synonymous, this review follows suit for simplicity. Importantly, however, leading researchers in the field advocate against equating these terms, as will be briefly elaborated in the following subsection.

7.1.3. Distinguishing unconscious from implicit

Within the context of unconscious bias, the term "unconscious" implies several distinct meanings [19]:

1. Absence of phenomenology: an individual lacks conscious experience or feeling associated with the biased mental state.
2. Unawareness of the content: a subject may be unable to identify the specific representations impacting their performance on implicit measures.
3. Unawareness of source and impact: an individual is unable to understand where the biases originate from or how they influence the actual behavior.
4. Unawareness of the relations: an individual does not recognize that their implicit and explicit evaluation of a given target differ.
5. Variation in awareness: an individual's insight into their own mental state is variable (e.g., having "access" awareness, but lacking "phenomenal" awareness).

Implicit biases were (and commonly still are) described as discriminatory biases that are grounded in implicit attitudes or implicit stereotypes [20], where "implicit" means automatic or unconscious. In this sense, there is plenty of overlap between the designations *unconscious* and *implicit*. However, over the last two decades, the term "implicit" (in the context of implicit bias) was redefined to mean "indirect," making it incompatible with the modifier "unconscious" [21]. There were several reasons for doing so. First, researchers sought to establish conceptual clarity, given the lack of consensus on how to conceptualize and operationalize terms such as "automatic" or "unconscious". Second, theorists recognized that while indirect methodologies uncover unconscious processes and attitudes, they might also reveal biases or preferences that are well-known to the subjects. Some studies have demonstrated that people can accurately gauge their implicit biases when specific methodological modifications are applied [22]. A related example is the "bogus pipeline" experiment [23], in which participants were led to believe that deceptive self-reports could be detected by the researcher. This modification resulted in a significantly stronger correlation between explicit self-reports and implicit scores than was typically observed under standard testing conditions. Together, these studies provide evidence that people are aware of the content of their implicit biases to some extent - or, as some research suggests [24], to a considerable extent. Although this awareness does not necessarily indicate direct introspective access to underlying mental content, as individuals may instead infer their biases from various other available cues [25].

In view of these considerations, the currently proposed definition of implicit bias [9] emphasizes its two features: 1) it is expressed and measured indirectly, and 2) it shapes biased judgment and discriminatory action that *can* operate outside of the person's awareness. As such, this formulation is at least partly incongruous with the definition of *unconscious* bias. It is worth noting, however, that there is no

consensus regarding the exact definition of "implicit". Moreover, there is substantial research that challenges the rigid dichotomy between implicit and explicit, or conscious and unconscious.

7.1.4. Beyond the implicit – explicit dichotomy

Favored by some scholars for its seemingly uncontroversial and theoretically uncommitted formulation, the strict definition of "implicit" means "indirectly measured," whereas "explicit" denotes "directly measured" [26]. The former relates to indirect measures of bias – most notably the Implicit Association Test – as well as physiologically based techniques, such as functional neuroimaging, startle eyeblink response, electromyography, and event-related potentials; the latter consists predominantly of explicit self-report surveys. In the broader scientific discourse, "implicit" is equated with "unconscious" – that which is inaccessible to awareness, and, by extension, leads to unintentional action. "Explicit" designates that which is conscious, accessible, and deliberate. These two modifiers, together with their associated meanings, are applied to bias-related concepts, such as stereotype, prejudice, and discrimination, to delineate two distinct facets of the same phenomenon (for definitions, see Table 1. Definition of concepts). Increasingly, however, research calls this dualistic framework into question [27, 28].

Table 1. Definition of concepts

Concept	Definition
Implicit (technical)	Indirectly measured
Explicit(technical)	Directly measured
Indirect measure	An assessment of a mental construct by observing automatic responses, assuming the individual has no conscious access to that construct
Direct measure	An assessment of a mental construct through explicit self-report, assuming the individual has conscious access to that construct
Implicit (general)	Mental content that is inaccessible to awareness and drives unintentional action
Explicit (general)	Mental content that is conscious, accessible, and deliberate
Attitude	A positive or negative evaluation of an object, person, group, or idea

Concept	Definition
Bias	A tendency to favor or disfavor a target in ways that may produce unfair or unequal judgment
Stereotype	A belief or an opinion about the characteristics and attributes of a social group and their members
Prejudice	A personal attitude directed toward social groups and their members
Discrimination	The behavioral manifestation of prejudice and stereotype, resulting in unequal treatment

For some time, it was assumed that indirect methodologies captured phenomena that differed substantially from explicitly gathered content. In other words, it was thought that individuals could hold two distinct evaluations of the same target (e.g., a negative implicit attitude alongside a positive explicit attitude toward cyclists, or an implicit stereotype associating marathon runners with boredom, while consciously regarding them as interesting and captivating people). Low correlations between implicitly and explicitly measured results were interpreted as indicating that implicit measures record unconscious content that is inaccessible to introspective self-report surveys. However, it is now widely recognized that a range of factors may contribute to this apparent divergence, including self-presentation concerns and socially desirable responding. In addition, recent extensive meta-analytic reviews show that implicit and explicit measures are, in fact, always positively correlated [26], thus indicating the inherent similarity between implicit and explicit contents [27].

Another point of contestation concerns the assumption that implicit bias remains rigid even in the face of new, contradictory evidence that is explicitly reinforced. Such a dynamic lends itself to the interpretation that implicit bias reflects highly overlearned information (e.g., old attitudes) that is resistant to change, while explicit bias refers to the recently learned and transient material. Considerable evidence for this perspective comes from a pivotal, high-quality study on implicit racial bias [29] with an astounding sample size that ran bias reduction interventions in university campuses and measured their effectiveness. It concluded that while there was a significant reduction in implicit bias shortly after the interventions, none of them had a lasting effect after 1 or 2 days. These findings were considered strong evidence for the rigidity of implicit biases. However, other studies have demonstrated the contrary: that implicit associations are labile, context-dependent, and temporally unstable. Research indicates that explicit biases are, in fact, more temporally stable than implicit biases [30]. While aggregate levels of

implicit bias across populations are highly stable, individual levels fluctuate significantly. It is well established that the same person can yield entirely different results on an Implicit Association Test within a few hours. Contextual cues can significantly impact these measurements [31]. For example, if a person who is generally against capital punishment has recently read an article advocating for the death penalty, the test will register a preference for the death penalty, even though on another occasion the same person would score in a vastly different way [27].

This distinction is further complicated by the fact that implicit measures are not "process-pure", but instead capture a range of cognitive and affective processes, including beliefs and explicit attitudes [19]. Thus, what "implicit" entails is not necessarily what one is unaware of, but also what one leaves unattended or unspoken in a given context. On the other hand, explicit expressions of subjective experience are active constructions [32] based on past experiences and may not involve conscious deliberation. Therefore, it is reasonable to conclude that what is expressed overtly may be the direct consequence of automatic associations, and vice versa, what appears inconspicuous and unintentional may result from concealed deliberation. This discussion is particularly relevant when examining implicit and explicit forms of discrimination. As such, the ambiguous boundary between the two justifies a broader and more inclusive evaluation of such phenomena.

7.1.5. Stereotype, prejudice, discrimination

At its core, bias operates through three psychological processes: cognitive, affective, and behavioral [33]. Stereotypes represent the cognitive component, referring to beliefs or opinions about the characteristics and attributes of social groups and their members. Prejudice represents the affective component, involving evaluative attitudes directed toward social groups and their members [34]. Discrimination is the behavioral manifestation of stereotypes and prejudice, leading to unequal treatment. These processes operate on a range from highly automatic to highly deliberate. As such, unconscious bias designates the automatic end of the spectrum. Therefore, any discussion of unconscious bias ultimately refers to the automatic activation of stereotypes or prejudice that may or may not lead to discrimination.

7.1.6. Relation with cognitive bias

Unconscious bias can be considered a subtype of cognitive bias. Although there is no uniform definition of "cognitive bias" [35], the concept is generally understood as automatic information processing that relies on mental shortcuts, or heuristics. Such processes guide decision-making through statistical

inferences rather than deliberative thinking. Thus, it is a process that uses minimal cognitive resources and has even been shown to be efficacious [36]. Nevertheless, cognitive biases are viewed negatively, and not without reason, as they occasionally lead to highly costly errors in such fields as medicine, the justice system, finance, and management [37]. Although many cognitive biases are fundamentally neutral, they provide the cognitive backbone that reinforces unconscious biases [38] and leads to the propagation of unintentional discrimination. Several common cognitive biases that operate within the medical field are presented in the following section.

The concept of cognitive bias (including implicit bias) is primarily evaluated through the framework of Dual-Process Theory [39]. Its central claims posit two modes of cognitive processing, commonly referred to as System 1 and System 2. System 1 is fast, efficient, intuitive, automatic, unconscious, associative, and dates to evolutionary past. Conversely, System 2 is a more recently evolved capacity for rule-based processing that is slower, more effortful, analytical, and associated with conscious deliberation. Under this paradigm, cognitive biases arise when System 1 relies on heuristics, utilizing metabolically efficient shortcuts that are often adaptive, though occasionally prone to error.

7.1.7. Relation with systemic bias

Closely related to unconscious bias is the concept of systemic bias. This refers to the way in which societal structures and institutional processes create or perpetuate inequality and mistreatment without deliberate intent [9]. Discriminatory practices are thus inherent to societal systems, independently of the attitudes and opinions of those who occupy roles within the system. This phenomenon has been aptly called “racism without racists” [40], and most likely applies to other forms of discrimination as well. People learn their attitudes and opinions from their surrounding culture, which is largely shaped by these societal structures. It is theorized, therefore, that the average level of implicit bias in a community reflects the amount of bias present in the system itself. Thus, implicit bias is an indicator of a given system’s level of bias [41]. In addition, implicit bias can also serve as a vehicle through which systemic prejudice is embodied and passed on. Evidently, it can be a mutually reinforcing process, where systemic bias provides the content of implicit bias, and the ensuing discrimination further solidifies existing societal structures.

Therefore, given that implicit biases originate from the surrounding cultural landscape, it is reasonable to view them as individual-level expressions of systemic bias. Moreover, while the broader cultural environment represents a static baseline of collective attitudes, the accessibility of stereotypical cues varies from one immediate context to another, which is reflected in the fluctuating individual

performance on implicit measures. From this perspective, it becomes apparent why the aggregate level of implicit bias across populations remains stable and why most interventions show no long-lasting efficacy, despite implicit bias being highly malleable on a personal basis. This is the core idea behind the "Bias of Crowds" framework [41].

7.2. Unconscious bias in health and healthcare

This section examines the impact of unconscious bias on health outcomes and explores how it manifests within the healthcare system.

7.2.1. Bias, stress, and allostatic load

When expressed as discrimination, unconscious bias has severe, detrimental effects on health outcomes [42]. The primary pathway involves the induction of chronic stress, which causes cumulative physiological damage over time. Such an effect on physical health is usually studied within the framework of allostatic load. Allostasis involves anticipating and meeting a body's physiological needs before they actually arise, which requires predictively adjusting internal set-points to match environmental demands. However, when these regulatory systems are repeatedly activated without adequate recovery, this results in a cumulative biological wear and tear called allostatic load [43]. Marginalized and stigmatized populations are, by definition, the primary targets of stereotypes and prejudice, bearing the greatest burden of discrimination. This marginalization occurs across lines of race, ethnicity, gender, sexual orientation, socioeconomic status, age, physical characteristics, and disability. It is not surprising, then, that stigmatized populations consistently display higher levels of allostatic load than the non-stigmatized groups [44].

A social environment fraught with discriminatory stressors can induce a persistent state of readiness for threats and adversities. In response, the physiological system becomes locked in sustained activation of the hypothalamic-pituitary-adrenal axis and the sympathetic-adreno-medullary axis, with this activation persisting even during normal periods of rest. This leads to prolonged elevations in cortisol, epinephrine, and norepinephrine levels, thereby causing allostatic load. Sustained periods of allostatic load initiate a cascade of secondary and tertiary physiological responses. In the cardiovascular system, this results in elevated resting blood pressure, significantly increasing the risk of coronary heart disease, ischemic heart disease, and peripheral arterial disease. Metabolically, allostatic load is strongly correlated with increased body mass index, elevated waist-to-hip ratios, and the earlier onset of metabolic syndrome. In addition,

it leads to elevated fasting glucose, cholesterol, and triglyceride levels, predisposes to insulin resistance, and contributes to the development of type 2 diabetes. Moreover, chronic exposure to perceived discrimination is associated with a higher incidence of mental illnesses, such as clinical depression, generalized anxiety disorder, and severe psychological distress [45]. Furthermore, the anticipation of bias leads those experiencing it to a state of hyper-vigilance, which strains cognitive resources and impairs executive control. As a result, the capacity for self-control becomes depleted [46,47]. This depletion increases the inclination toward unhealthy coping mechanisms, such as smoking or substance abuse, which temporarily attenuate distress [48], or, conversely, decreases the likelihood of maintaining effort-requiring healthy behaviors, including treatment adherence and participation in preventive protocols. These behavioral responses create further avenues for discrimination down the line. This demonstrates how chronic subjection to overt or covert discrimination initiates a self-reinforcing cycle [49], wherein health and life satisfaction steadily deteriorate.

7.2.2. Unconscious bias in clinical care

Clinical care is a field especially sensitive to the impact of bias and discriminatory behavior. Biases affect patient care and clinical training, obstruct communication between patients and healthcare professionals, as well as among colleagues, and influence hiring, career advancement, and allocation of research funding.

Healthcare professionals are highly susceptible to various cognitive biases in their decision-making [50]. While most of them offer clear advantages under the time and resource constraints of routine medical practice – especially in emergency departments – they occasionally lead to costly, preventable errors [51]. Some of the important heuristics operational in the everyday clinical environment are listed in Table 2. Definitions and examples of cognitive biases [52,53]While these examples demonstrate how cognitive biases impact clinical decision-making under standard conditions, this influence becomes even more critical when combined with a clinician's unconscious biases.

Table 2. Definitions and examples of cognitive biases [52,53]

Type of bias	Definition and examples
Anchoring bias	Being disproportionately influenced by the first piece of information encountered and making judgments accordingly, even when later information contradicts the first impression.

Type of bias	Definition and examples
	A patient presenting to an emergency room with acute shortness of breath and a history of asthma leads the clinician to assume that the cause of the symptoms is an asthma attack, and after administering bronchodilators with no effect, the clinician remains fixated on the asthma diagnosis rather than considering alternatives such as pulmonary embolism.
Availability bias	Overestimating the likelihood of an event based on the recent, memorable, or emotionally charged experience. Suspecting a very rare condition in a patient with vague symptoms after reading about it in a scientific article a day before.
Base rate neglect	Prioritizing a specific instance instead of considering the general prevalence of an event. Concluding that a COVID-19 vaccine is ineffective because vaccinated patients have a higher rate of hospitalization than the unvaccinated, while ignoring that a vast majority of the population is vaccinated – a base rate that makes such an outcome statistically predictable. Diagnosing a patient with a very rare illness instead of considering the possibility of a false-positive answer. Assuming a young patient with chest pain has a rare cardiac disorder while overlooking the much higher base-rate likelihood of benign musculoskeletal pain.
Commission bias	Prioritizing doing something instead of nothing, even when the evidence suggests that action may lead to worse outcomes. Prescribing infusion therapy to a palliative care patient when there are no clinical indications to justify the intervention. Prescribing antibiotics for a likely viral upper respiratory infection because doing something feels preferable to idly waiting.
Confirmation bias	Seeking or favoring information that supports existing beliefs while overlooking contradictory evidence.

Type of bias	Definition and examples
	Attributing a psychiatric patient's anger to the worsening of his condition, failing to consider situational factors, such as resentment over their involuntary commitment
Conjunction fallacy	Assuming that a combination of events is more probable than a single event. Concluding that a patient presenting with acute shortness of breath and chest pain is more likely to have both a pulmonary embolism and deep vein thrombosis, rather than just a pulmonary embolism, despite the single diagnosis being mathematically more probable.
Diagnostic momentum	Accepting a preliminary diagnosis that gets established and passed on among clinicians without sufficient critical review, making it harder to revert. Continuing to treat a patient according to the primary diagnosis of a gastrointestinal infection made by a night-shift physician without reconsidering alternative explanations, such as appendicitis.
Overconfidence	Maintaining high certainty in a judgment without adequately evaluating its factual accuracy. A doctor confidently diagnosing a viral upper respiratory tract infection without adequately assessing the validity of their opinion (especially problematic with physical examinations which are prone to inaccuracy, such as auscultation).
Representative bias	Making a judgment about an event or subject based on the degree to which it resembles a typical prototype. Dismissing the possibility of a pulmonary embolism in a young, athletic person complaining of shortness of breath only because the patient does not match the prototypical demographic.
Search satisfying	Ceasing the search for further or alternative information after an initial, satisfactory explanation is found. A radiologist, after identifying an obvious fracture on an X-ray, halts their review, missing a second, more subtle fracture.

Type of bias	Definition and examples
The framing effect	Reacting to a choice differently depending on how the information is presented. A surgeon is significantly more likely to obtain patient consent when describing a procedure as having a 90% survival (or success) rate rather than a 10% mortality (or failure) rate.

At the patient level, individuals from marginalized groups face the greatest risk of being subjected to unconscious biases. While prejudice toward these groups may be rejected at a conscious level, ingrained cultural associations continue to inform implicit attitudes, thereby impacting treatment outcomes. For this reason, healthcare providers must remain mindful when interacting with these populations to ensure they deliver optimal care. Drawing on a review article [54], the following table (Table 3. Bias-related disparities in medical care according to marginalized and stigmatized groups) outlines clinical areas where diagnostic biases commonly occur across different populations.

Table 3. Bias-related disparities in medical care according to marginalized and stigmatized groups

Marginalized and stigmatized group	Bias-related disparities in medical care
Women	Dismissal of symptoms as psychosomatic or minimization of symptom reports [55]; delayed cardiac intervention due to misattribution of symptoms to non-cardiac causes [56,57]; systematic lack of standardization regarding female-specific physiology in medical research and practice, leading to diminished treatment efficacy [58]; significantly poorer surgical outcomes when treated by male surgeons [59]
Racial and ethnic minorities	Under-assessment of pain levels and delayed analgesic administration [60]; lower rates of thrombolysis, cardiac catheterization, and noninvasive diagnostic testing [61]; late diagnostic referrals and reduced access to preventative cancer screening programs [62]; diminished quality of care due to frequent antagonizing interactions between clinicians and patients [63]

Marginalized and stigmatized group	Bias-related disparities in medical care
People with mental illness	Psychiatric history as a justification for lower-quality medical care [64]; dismissal of physical complaints, particularly pain, leading to the neglect or delayed recognition of health problems [65]
Older adults	Reduced access to and provision of mental health treatment compared to younger populations [66]
People living with HIV	Hindered expansion of public health measures for prophylaxis, diagnosis, and treatment [67]; reduced rates of HIV testing, lower treatment adherence [68]
People from the LGBTQIA+ community	Reduced access to specialized and gender-affirming care; lower quality of care and poorer health outcomes associated with stigma, discrimination, and avoidance of healthcare due to anticipated mistreatment [69]
People with substance use disorders or histories of incarceration	Diminished feelings of empowerment and significantly poorer treatment outcomes [70]; unfair clinical judgment of "treatability" prior to admission, which can restrict or delay access to essential rehabilitation programs [71]
People with low socioeconomic status and people experiencing homelessness	Suboptimal access to and delivery of healthcare services, along with higher rates of hospitalization associated with delayed intervention, structural barriers, and systemic neglect [72,73]
People with disabilities	Suboptimal preventive care and poorer overall health outcomes [74,75]; lack of clinical preparedness and insufficient education of clinicians regarding the unique health needs of people with disabilities [76]

However, it must be emphasized that properly addressing unconscious bias is challenging for healthcare professionals, given their taxing working conditions - high patient loads, difficult long shifts, extensive bureaucratic responsibilities, and significant mental strain. While individual efforts, such as practicing

mindfulness [77] or maintaining a compassionate outlook, can certainly improve the working environment [78], structural change is essential. Creating safe environments for reflection on experiences with implicit bias, reducing the pressure to make important decisions under high stress, and providing role-playing opportunities with the patients when unconscious bias is encountered [79] contribute to reducing the impact of implicit bias on health outcomes [54].

The negative impact of unconscious bias on health is indisputable, to say nothing of its pervasive influence across other spheres of life. This phenomenon, therefore, clearly warrants thorough investigation. Its universality across human populations has led scholars to posit evolutionary foundations for these biases [80 – 82].

7.3. Evolutionary perspectives on unconscious bias

This section introduces the core principles of evolutionary theory and outlines the foundational assumptions of this approach. It then examines several prominent evolutionary perspectives on unconscious bias, followed by a critical evaluation of these theories and the presentation of an alternative domain-general framework in the subsequent two sections. Although these principles apply to a wide range of phenomena, this analysis focuses specifically on unconscious bias.

7.3.1. Principles of evolutionary theory

Evolution is generally understood as a process by which biological populations accumulate modifications over successive generations. These modifications are encoded in DNA and passed on to offspring as genes via sexual reproduction cells, called gametes. Genes condition specific phenotypic expressions of an organism: its anatomical constitution, physiological states, and behavioral patterns, or simply, its traits. If a gene influences a particular phenotype, it is regarded as a gene for that phenotype. A single gene can have alternative forms called alleles, each coding for a specific variation of the phenotype. The total combination of alleles constitutes the organism's genotype. Importantly, phenotypic modifications acquired during an organism's lifespan do not get passed on to offspring, as they are not encoded in DNA; only genetic information is transferred across generations. In short, the genotype influences the phenotype, but the phenotype doesn't affect the genotype [1]. As such, genes constitute the cornerstone of

[1] This one-way relationship is not wholly accurate. Empirical evidence demonstrates that organisms can acquire phenotypic modifications that are transmitted to offspring through non-genetic pathways, as illustrated by transgenerational epigenetic inheritance [83]. It is also worth noting here that the classic framework of evolutionary science is based on the Modern

evolutionary processes, and contemporary evolutionary biology defines evolution primarily as *an alteration* of gene and genotype frequencies down a line of descent [86]. The other essential piece to consider is what causes this alteration. However, before exploring this second part, one caveat regarding the relationship between genotype and phenotype needs to be addressed.

Genes do not strictly determine phenotypes; rather, phenotypic expression emerges from the interaction between genes and the environment. The same genotype can produce different traits if developmental environments vary sufficiently[2]. For example, two tomato plants grown from identical seeds but planted in different soils – one arid and the other nutrient-rich – will reach different heights and yield qualitatively different fruits, despite having the same genotypes. Following this logic, a *gene for* a particular phenotype will produce *that* phenotype only under some developmentally relevant conditions, and not others. From this, two implications emerge [86]:

1. An increase in a gene's frequency leads to a corresponding increase in the associated phenotype's expression *only when* the relevant developmental conditions stay relatively stable across generations. The environment could change to such an extent that it would no longer support the expected phenotype, and then an increase in gene frequency would not result in an analogous increase in phenotype expression. Therefore, relative environmental stability is a prerequisite for the consistent expression of expected phenotypes via genetic inheritance.
2. Phenotypic variation can occur solely due to changes in the environment, without any relevant alterations in genotypes. Buller refers to these exclusively environmentally driven shifts in phenotypic frequencies across generations as *phenotypic evolution* [86, p26]. This process is clearly illustrated by the significant increase in the average height of human populations in the twentieth century due to widely improved living conditions [87].

Evolution is driven by two main causes [86]: changes *within genes*, which add entirely novel traits into the gene pool, and changes in the overall *frequency of genes* over time. Regarding the first process, new traits emerge through two distinct mechanisms within reproductive cells: mutation, which creates *new genes* via random changes to the DNA sequence, and recombination, which randomly shuffles existing

---

Synthesis [84], and there have been numerous innovations within adjacent fields that challenge this framework and call for its extension [85].

[2] Certain genotypes are highly resistant to environmental influences and consistently produce identical traits (such as genotypes for Huntington's disease or cystic fibrosis). However, such direct correspondence between genotype and phenotype is relatively rare and represents an exception rather than the rule.

genetic variants into *new combinations*. The second cause of evolution - changes to the frequencies of genes - also occurs through two processes: genetic drift and natural selection [86]. Genetic drift describes the random change in allele frequency resulting from random sampling of reproductive cells and chance survival [3] due to catastrophic events. In these cases, allele frequencies fluctuate strictly by chance, completely independent of any phenotypic advantage. Natural selection, by contrast, is highly reliant on phenotypic advantages and is considered the primary driver of evolution in large populations, including humans. It occurs when there is *phenotypic variation* within a population, when these phenotypes are *genetically heritable*, and when these variants confer different *fitness* advantages [86].

Fitness is a central concept in natural selection because it reflects an organism's ability to survive and reproduce, which dictates its chances of successfully transmitting its genes to the next generation. An organism's fitness is determined by how well its traits adapt to the specific environmental challenges, or *selection pressures*. The more fit an organism is, the better equipped it is to solve adaptive problems posed by its environment. As a result, a fitter organism will, on average, contribute more to the gene pool of its population. Thus, natural selection increases the prevalence of genetic variants that aid survival and reproduction, diminishes those linked to disadvantageous traits, and exerts no selective pressure on neutral variants [86].

When a phenotype confers an adaptive advantage and increases in frequency across generations, it is said that there has been *selection for* that phenotype. Selection acts only on genetically heritable traits. Any trait that arises in a population through a history of selection is termed an *adaptation*. Crucially, a trait can be adaptive without being an adaptation; likewise, an adaptation may no longer be adaptive. This is because *adaptiveness* refers solely to the *current fitness* of a trait, which may not have evolved under natural selection and thus would not qualify as an adaptation. Conversely, an adaptation may lose its adaptive value when ecological pressures change [86].

With these concepts established, it can now be illustrated how natural selection leads to traits becoming universally prevalent, or fixated, in a population. The process begins when a mutation or genetic recombination improves an existing structure, conferring a fitness advantage. Individuals carrying this adaptive trait achieve higher reproductive success, causing the trait to increase in frequency in subsequent generations until it reaches fixation. Whether unconscious bias evolved in a similar way as direct adaptation, or is merely a by-product of other general processes, remains to be answered.

[3] As an illustration of chance survival, if a catastrophic event indiscriminately eliminates a large portion of a population carrying a specific allele, the survivors are left with an alternative allele at a much higher frequency. As a result, this alternative allele would increase in frequency the next generation, thus constituting evolution, albeit driven by chance.

7.3.2. Introduction to an evolutionary approach

Evolutionary approaches operate under a set of core assumptions [80]. A foundational claim upholds that the human brain has evolved through natural selection. This means that the brain has been molded by various selection pressures affecting human populations in our evolutionary past. Those who possessed relevant adaptations had better survival and reproduction rates, and thus their adaptations propagated from one generation to the next. In this way, contemporary human populations came to have the propensity for stereotyping, prejudice, and discrimination. However, what was adaptive in ancestral environments may no longer be adaptive in the current one; despite this, these ancient adaptations continue to shape human psychology to an extent.

Broadly speaking, evolutionary reasoning used to infer adaptations adheres to a specific logical structure consisting of three parts [86]. First, the ancestral environment posed adaptive problems to early humans. Second, adaptive solutions arose in response to those problems driven by genetic change. Third, these adaptations were sufficiently successful that they were selected to fixation and now constitute a universal human nature [88]. In evolutionary psychology, this reasoning relies on a method called *evolutionary functional analysis* [86], which is characterized as the reverse-engineering of the human mind from the vantage point of the Pleistocene epoch [4]. In practice, the first step involves theorizing historical selection pressures universal to ancient human populations. The second step requires inferring the psychological mechanisms that developed as adaptations to these pressures. The third step entails conducting psychological experiments with contemporary populations to determine whether these mechanisms actually exist. For heuristic purposes, the first step is referred to as *adaptive problems*, and the second step as *adaptive solutions* to these problems. With the core assumptions and structure of evolutionary reasoning established, the following subsection delineates several evolutionary perspectives on unconscious bias.

7.3.3. Origins of unconscious bias

Ultimately, the primary adaptive problem the human brain evolved to solve is maximizing reproductive fitness, which entails successful reproduction and the propagation of one's genes into future generations [80]. To achieve this, beyond the obvious (securing a mate), many auxiliary challenges must be addressed.

[4] Evolutionary psychologists view the Pleistocene epoch (from 2.58 million to 11,700 years ago [89]) as the temporal backdrop for the Environment of Evolutionary Adaptedness: a set of historical selection pressures under which human psychological adaptations reached fixation.

Foremost, humans must survive long enough to reproduce, which requires securing sustenance and avoiding harm from predators, diseases, or other humans. Furthermore, they need to ensure their offspring mature to reproductive age. Moreover, they have to navigate a broad range of social dynamics to enhance their likelihood of reproductive success, such as attaining high social standing, cooperating with others for difficult tasks, and guarding partners against rivals. The social domain is particularly important, as above all, humans are a highly interdependent species [90]. Social interactions can offer opportunities that lead to significant adaptive benefits. Other times, they can be severely damaging, inflicting substantial fitness costs. As such, individuals must effectively distinguish between favorable social situations worth pursuing and potentially threatening ones that would be better avoided. Within this context, some researchers [80] theorize that biases evolved as *affordance management systems*: specialized psychological mechanisms designed to navigate the contingencies of social situations. Accordingly, stereotypes, prejudices, and discrimination are envisioned as some of the instruments through which these systems operate. The rationale of this perspective can be succinctly illustrated in two parts: (1) In ancestral environments, certain statistical regularities existed between individuals' observable traits and their affordance-relevant behaviors; and (2) Ancestors who recognized these patterns made more effective judgments than those who did not, thereby gaining substantial survival and reproductive advantages. As before, the first premise outlines the adaptive problem, while the second represents the adaptive solution. The following accounts apply this bipartite evolutionary reasoning to explain how particular biases arose across the dimensions of age, sex, race, ecology, disease, coalition, and kinship. Each dimension is examined independently, drawing directly on the work of Sng and colleagues [80].

7.3.3.1. Age

An important theoretical backbone is provided by life history theory [91], according to which organisms have a limited energy capacity; therefore, they must effectively allocate resources across various tasks throughout their lifespan to achieve evolutionary success. Some tasks, such as securing a mate and raising offspring, are only relevant during specific life stages. As such, interacting with everyone indiscriminately is inefficient because individuals' goals and physical capacities differ by age. Categorizations based on age are adaptive because they aid in gauging others' behaviors and abilities [80]. For example, age-based cues allow for the instant distinction between a child requiring care and an older adult offering wisdom. Similarly, they help in identifying vigorous young adults, which carries both competitive risks and cooperative benefits, or those of reproductive age who may qualify as viable partners.

7.3.3.2. Sex

Similar to categorizations by age, grouping people by biological sex is useful as it provides clues about their social goals and physical capabilities. Females invest significant energy in reproduction, from pregnancy to nursing, and therefore tend to be more selective in choosing mates. Males, on the other hand, are pressured to compete for female attention. As a result, identifying a person's sex allows for the rapid assessment of social encounters, such as recognizing a quality mate or assessing the risk of conflict with a rival. Additionally, the affordance implications of age and sex are usually intertwined, as this combination provides highly relevant information. For example, women can only have children during certain years of their lives, and men are typically most competitive as young adults. In sum, age and sex provide important cues regarding others' abilities and behaviors, and recognizing these cues informs individuals about social opportunities and threats. For this reason, it is likely that humans evolved a propensity for bias based on age and sex [80].

Another component of sexual bias concerns prejudices toward homosexuality, which are primarily experienced in the form of disgust. Since disgust evolved as a protection mechanism against infectious disease, heterosexual individuals with sexual prejudice may implicitly react to same-sex non-heterosexuals as if they were physically contagious. Such reactions could serve as a defense against the potential costs of unwanted sexual attention [80].

7.3.3.3. Race

The evolutionary origins of race-based bias are not straightforward. Early humans lived in relatively homogeneous groups and rarely encountered individuals from distant, racially different populations, making it highly improbable that categorization based on race would develop. In light of this, two major interpretations view race as a proxy for highly relevant social information: one's ecological circumstances [92] or coalitional belonging [93,94].

To elaborate on the former, environmental conditions determine which behavioral strategies are more frequently encountered. In unstable environments that lack resources and are highly perilous, it is more adaptive to assume a "fast" strategy focused on immediate rewards. This strategy involves various behavioral traits, ranging from earlier sexual initiation and more frequent sexual activity with multiple partners to heightened impulsivity and an increased proneness to criminal behaviors [95]. Conversely, environments associated with prosperity and high living conditions lead to the adoption of "slow" strategies that are future-oriented and characterized by behavioral traits opposite to those of "fast" strategies. Taking all of this together, correctly anticipating others' ecological backgrounds has significant affordance implications. The ability to categorize people according to their inferred ecological

circumstances, and thus predict their behavioral inclinations, provided fitness advantages to ancestral populations [80]. In contemporary societies, due to the contingencies of historical and postcolonial relations, marginalized regions experience disproportionate hardship and exploitation compared to dominant economic hubs [96]. The same dynamic repeats on a local level, where marginalized racial groups are subjected to systemic disadvantages compared to majority populations, causing their respective ecological niches to differ drastically. As a result, race becomes associated with specific environments, and this ecological disparity then serves as a basis for racial categorization and discrimination [80].

The other interpretation considers race as a proxy for categorizing individuals based on group membership and coalitional affiliation. The argument unfolds like this: early humans belonged to groups and coalitions for adaptive purposes. Within these coalitions, there was a general uniformity in how people looked; conversely, distinct physical features indicated that individuals belonged to a separate coalition. To survive, people had to quickly distinguish with whom they could cooperate and who posed a potential danger. As such, there was significant selection pressure to learn to differentiate others according to their physical appearance. Because race is a highly salient visual cue, it became a target of this coalition tracking mechanism, thus explaining the prevalence of racial bias [80].

7.3.3.4. Disease

Pathogens have constantly threatened human survival, meaning the ability to identify and avoid contagious individuals is highly adaptive. However, developing a cognitive mechanism capable of accurate detection of infectiousness was improbable for ancestral populations, given the variety of constantly evolving microorganisms and the highly variable, inconsistent symptoms they produce. Instead, an alternative adaptation emerged that uses various phenotypic irregularities, such as bodily disfigurements or behavioral abnormalities, as markers of latent or active infections. This adaptive solution has been aptly called the behavioral immune system [97]. It operates by broadly scanning others for any atypical physical or behavioral traits, implicitly categorizing anyone with these traits as potentially diseased and best avoided. However, given its crude algorithm, it not infrequently detects false positives, interpreting healthy individuals with unusual appearances as potentially contagious. In this way, people who are obese, elderly, physically disabled, or living with a rare genetic condition become the targets of unwarranted discrimination. Despite these frequent misattributions, this mechanism is grounded in the robust evolutionary logic of error management [98]. Failing to detect a highly contagious person could be fatal, while mistakenly avoiding a perfectly healthy but unusual-looking person carries comparatively little biological cost. This principle is vividly illustrated by the

smoke detector metaphor [99]: it might trigger when no threat is present, but it will definitely go off when a real fire starts. Thus, it is hypothesized that many psychological adaptations, especially those concerning threat detection, are wired to be overreactive rather than risk missing a real danger [80].

7.3.3.5. Coalition

As previously noted, early humans relied on small cooperative coalitions for survival. These coalitions frequently clashed with rival factions, making intergroup hostility a persistent evolutionary threat. Moreover, intense rivalry over resources and sexual partners caused sub-groups to form, leading to violent internal disputes. Such a hostile nature of ancestral environments is supported by evidence that nonhuman primates and present-day foragers frequently engage in lethal confrontations. Under these circumstances, successfully separating allies from enemies became a survival necessity, driving a selection pressure for group-based categorization. The psychological mechanism, which once helped assess potential aggression or cooperation based on perceived coalitional affiliation, now underlies widespread discrimination based on ethnicity, nationality, religion, and other social identities. The male warrior hypothesis [100] suggests that ancestral males experienced additional evolutionary pressures to distinguish the outgroup males than females did. To secure fitness advantages given the fiercer competition for sexual partners, males more frequently formed alliances and engaged in group violence. For this reason, male psychology evolved specialized mechanisms for fighting rival factions. As a consequence, modern men are more vigilant to outside threats and experience higher hostility toward out-group men than women do [80].

7.3.3.6. Kinship

Individuals who prioritized caring for their direct kin ensured a higher probability that their shared genes were passed on. To achieve this, they had to successfully distinguish between family members and genetically unrelated people. Because humans do not automatically know who their relatives are, a psychological mechanism that aided in kin recognition was selected for. In this way, modern humans came to possess a relatedness-detection system, which registers environmental cues consistently associated with kinship, such as facial similarity, physical proximity, and shared outlooks. The ensuing categorization into the family group directly influences social behaviors and cooperative choices [80].

7.3.3.7. Synthesis and implications

Several themes apply uniformly across these dimensions of bias. To begin, each category is highly salient and carries clear affordance implications, thus explaining the selection pressures for developing adaptive solutions to navigate these different spheres [80]. The evolved psychological adaptations were selected

to fixation and are now universal in the human population. These adaptations are distinct, highly specialized brain modules tasked with dealing with their specific domains [101]. They operate automatically, mostly beyond conscious recognition, thus forming the basis of unconscious bias [80]. Stereotypes provide the cognitive content of bias, rapidly informing individuals about what to expect from social situations. Prejudices form the affective component of bias, triggering behavioral reactions in response to perceptions of social contingencies. Together, stereotypes and prejudices lead to discriminatory behaviors that can be either deliberate and highly directed or unintentional and performed unwittingly. Although these processes are often inaccurate, they reflect kernels of truth that aid in judging whether those encountered offer potential for harm or cooperation [80]. After all, according to error management theory, biases are calibrated like a smoke detector: it is safer to mistakenly assume potential dangers and miss out on opportunities than to overlook a genuine, life-threatening danger. However, being rigidly biased towards the worst also carries fitness costs, as expending energy to address false threats is wasteful. Therefore, some flexibility within bias mechanisms is expected. This flexibility varies depending on environmental circumstances, where individuals are open to alternative interpretations when feeling generally safe, and rely on biases when feeling threatened [80].

7.4. Critique of evolutionary approaches

The previous section recounted major evolutionary explanations regarding the origins of unconscious bias. While these accounts may appear well-founded, their validity remains highly contested. By drawing on the substantial critique advanced by philosopher of science David Buller [86], this section demonstrates that the premises underlying such explanations lack reliable evidence. To contextualize this critique, the initial subsection describes the teleological logic underpinning evolutionary research and illustrates how it aligns with evolutionary psychology's primary method of inquiry: evolutionary functional analysis. The core arguments are then presented along the three axes of this method: adaptive problems, adaptive solutions, and modern empirical evidence [86].

7.4.1. Teleology

Evolutionary research often exhibits a problematic tendency to ascribe intrinsic purpose or goal-directed intent to biological processes. In the philosophy of science, this is known as teleology: explaining the cause of a phenomenon by its goal or functional utility. Teleology underpins adaptationism, which assumes that unconscious biases were selected during evolution for distinct survival advantages. Such

teleological arguments rely on mental inferences about evolutionary goals rather than explicating the psychological mechanisms that actually drive biased responses. To be clear, there is nothing inherently wrong with teleological claims. They align well with common sense and are useful for generating hypotheses. However, teleological inferences cannot be treated as factual statements on their own, because they rely on metaphorical language that lacks empirical verification and should therefore be avoided when assessing the adaptiveness of biological features [102]. Nevertheless, they are frequently used to postulate causal relations, thus conflicting with established principles in evolutionary research [103].

To maintain scientific rigor, explanations of adaptive features should be grounded at a mechanistic level. In the terminology of Action Identification Theory [104], this corresponds to "low-level identification," where an action is expressed in its physical, mechanical, and concrete details. This stands in contrast to "high-level identification," which involves representing an action in terms of its broader meaning, purpose, or ultimate goals (see Table 4. Differences between mental inference and action identification level explanations). These distinct levels of explanation appear across various academic disciplines.

Table 4. Differences between mental inference and action identification level explanations[5]

Mental inference	**Action identification**
The rat freezes in fear	The rat has ceased movement due to motor-cortex inhibition
A human is angry	A human is scowling, shouting, and their blood pressure is rising
Unconscious bias exists as a threat detector that evolved as a psychological mechanism to quickly identify potential adversity	Unconscious bias involves the brain's rapid categorization of visual stimuli, which triggers an immediate physical arousal response in the nervous system

In philosophical tradition, for example, this distinction is captured by the terms 'immanent' and 'transcendent'. To illustrate, in literary theory, methods for analyzing a text depend on whether meaning is presumed to reside within the work (immanent analysis) or outside of it (transcendent analysis). An

[5] Barrett [102] refers to low-level identification as "action identification" and categorizes high-level identification as "mental inference." This review adapts this terminology.

immanent approach focuses exclusively on internal constitutive elements of a text, such as syntax and plot structure, disregarding externally derived concepts [105], and maps onto the "action identification" level of explanation. In contrast, the transcendent method interprets text through principles and contexts external to the text itself, situating it within the domain of mental inference.

In evolutionary sciences, the two primary levels of explanation rely on the distinction between the physical mechanism (the *proximate cause*, the "how") and the evolutionary purpose (the *ultimate cause*, the "why") [106]. The former approach focuses on the mechanical details of an organism's physical movements, physiological changes, and the developmental processes that produced them, coinciding with 'low-level' action identification. In contrast, the latter approach explores which adaptive factors could have contributed to a lineage's survival in the ancestral past, corresponding to "high-level" mental inference. Evolutionary functional analysis attempts to bridge these two levels of explanation: it infers the proximate causes of behavior (psychological mechanisms) from its ultimate causes (historical selection pressures). A reminder of how it works is worthwhile, as it will serve as a reference point for evaluating the validity of evolutionary psychological accounts of unconscious bias. The method begins by postulating the ancestral environments that exerted specific selection pressures on early humans. Next, it hypothesizes which behaviors, driven by psychological mechanisms, were adaptive and subsequently selected to near fixation given these pressures. Finally, it utilizes psychological experiments across diverse contemporary populations to confirm the existence of these inferred mechanisms. It will become evident that serious methodological flaws pervade every step of this method [86].

7.4.2. Adaptive problems

The Environment of Evolutionary Adaptedness constitutes the specific selection pressures under which human psychological adaptations evolved. Ancestral populations are assumed to have faced various adaptive problems, including mate selection, intrasexual competition, pathogen avoidance, and coalitional violence. By identifying these selection pressures, inferences can be made about the adaptations that evolved in response. To reconstruct adaptive problems, evolutionary psychologists rely on three sources of evidence: (1) the apparent design of psychological adaptations, observations of contemporary foraging societies, and comparative primate studies [86].

The first source of evidence involves inferring these pressures from the current design of psychological adaptations. However, this is problematic for two reasons [86]. First, this approach relies on circular logic: it uses current psychological mechanisms to postulate the historical selection pressures, which are then used to explain those very same mechanisms. Second, there is currently no definitive scientific

evidence that establishes which psychological features are genuine adaptations; thus, these features lack empirical grounding and cannot be used to reliably identify historical selection pressures.

The second source draws on studies of present-day foragers. These populations are assumed to maintain a lifestyle unchanged since the Pleistocene, offering a direct perspective on ancestral adaptive problems. This premise fails in two ways [86]. First, the assumption that contemporary foraging populations have been unaffected by the rest of civilization is simply not true. Anthropologists point out that human contact has long shaped these populations, impacting their livelihoods in considerable ways. As such, they cannot provide a privileged window into the ancient past. Second, these foraging societies are vastly diverse: some differ in mating and marriage systems; in some, women provide the majority of nutrition, while in others, their contribution is minimal. Because these structural differences significantly alter the adaptive problems of a given community, it remains unclear which population should serve as the proxy for ancestral populations, meaning that specific adaptive problems cannot be reliably inferred from living hunter-gatherer communities [86].

The third source lies in nonhuman primate studies. Since humans share a common lineage with other primates, it is suggested that the adaptive challenges present in these primate populations can shed light on the adaptive problems of ancestral human populations. However, this doesn't amount to reliable evidence as well [86]. The first reason is that the vast temporal distance separating humans from other primates makes it highly improbable that their Pleistocene adaptive problems were comparable. Humans and their closest relative, the chimpanzee, diverged from a common ancestor 5 to 7 million years ago. In the millions of years leading up to the Pleistocene, human and nonhuman lineages were exposed to vastly different environmental pressures as they occupied distinct ecological niches. So when early humans emerged in the Pleistocene, the adaptive challenges they faced were already entirely different from those of other primates. In addition, nonhuman primates have continued to evolve up to the present day, making their contemporary adaptive problems even further removed from those faced by early humans. The second problem resides in the difficulty of assessing which nonhuman primate is the closest candidate to represent human ancestors, as each primate species exhibits vastly different behaviors across social domains. Research shows that behavioral similarities among any species are best accounted for by similarities in their *ecological conditions*, not by morphological relatedness. Then, understanding how early humans behaved requires finding a modern primate that inhabits an ecological niche most closely paralleling that of early humans. However, the specific ecological niche early humans inhabited remains unknown. Without prior knowledge of the early human environment, a reliable primate candidate cannot be determined, and the adaptive problems cannot be reliably inferred [86].

Taken together, these points demonstrate that serious problems pervade the ability to identify empirically backed adaptive problems of ancestral human populations. Nevertheless, evolutionary psychologists argue that certain adaptive problems are undisputed and easily derived from the issues relating to survival and reproduction, such as mate selection, predator evasion, and nutrient acquisition. However, *broad* descriptions provide insufficient detail for determining the *specific* pressures that acted on the development of *particular* psychological adaptations [86]. Knowing a broad adaptive problem does not inform what an organism *must* specifically *do*. Every mammal faces these broad problems, but each has developed unique responses influenced by its specific environmental circumstances. An organism's traits dictate which ecological conditions actually impact its evolution (e.g., tall trees exert selection pressure only on species with grasping hands, whereas burrowing species remain unaffected). Determining the specific pressures that shaped early human psychology requires knowledge of ancestral cognitive and physical capacities. However, prior knowledge of their psychological structures is unavailable. Without it, hypotheses regarding adaptive solutions cannot be reliably established [86].

Finally, evolutionary functional analysis is predicated on the assumption that the adaptive problems faced by ancestral populations were relatively stable, allowing for the gradual development of specialized psychological adaptations. However, this assumption is also flawed for two reasons [86]. First, the adaptive problems relevant to the evolution of human psychology are primarily determined by *human psychology* itself. In other words, the development of psychological adaptations was driven by particular environmental pressures that were themselves mostly shaped by psychosocial aspects. To illustrate, if a cognitive trait were to evolve, it would impose new selection pressures on a community, in response to which perhaps another trait developed, once again impacting selection pressures, a process that would continue to reiterate, culminating in an "evolutionary arms race [107]. This makes such adaptive problems far from stable.

Second, humans actively construct niches that rapidly alter selection pressures. Just as dam-building beavers radically reshape their evolutionary environment, humans do the same in a wide variety of ways. Through agricultural and foraging practices, food preparation and preservation techniques, clothing and dwelling innovations, and complex social institutions, humans significantly alter the selection pressures affecting their populations. Once again, reconstructing specific selection pressures that shaped human psychology is hardly possible, as they shifted continuously according to human intervention [86].

However, even if it were possible to establish with certainty the adaptive problems of ancestral populations, difficulties in inferring adaptive solutions would still exist [86]. An adaptive problem may warrant a specific solution, but that does not necessarily mean that such a trait develops. As Buller illustrates, having night vision would have provided a significant advantage during the Pleistocene, yet

the trait never emerged. This is important to emphasize, because evolutionary reasoning often functions by assuming that if a feature would have provided an adaptive advantage, it most likely evolved. But there is no guarantee that this will happen. For an adaptation to evolve, a relevant genetic change must first occur. And usually, populations can survive without addressing particular adaptive problems [86]. Furthermore, inferring a solution from an adaptive problem is complicated by the fact that new adaptations always build upon previous structures. Consequently, it is unclear whether a particular feature arose in response to a specific adaptive problem or if it already existed. Without knowing these pre-existing structures, accurately identifying the contributions of specific solutions is highly improbable [86]. And as has been shown throughout this section, ancestral psychological structures remain unknown. This subsection has consistently demonstrated that:

1. Establishing a robust empirical basis for the adaptive problems faced by early humans is virtually impossible, as none of the proposed sources of evidence yield reliable data.
2. Broad assumptions about fundamental adaptive problems provide little utility in determining the specific selection pressures that shaped particular psychological adaptations; these pressures cannot be known, as their identification requires knowledge of pre-existing psychological structures.
3. There were no stable adaptive problems affecting psychological evolution, making it highly improbable that specialized psychological adaptations developed.
4. Adaptive solutions do not necessarily emerge in response to adaptive problems; even when they do, reliably identifying the specific traits that evolved remains difficult.

Without appropriate empirical grounding, postulations about the ultimate causes of human behavior become speculations based on mental inferences, and can be rightly called “just-so stories” [108]. The inability to establish a reliable basis for adaptive problems seriously calls into question the validity of inferences about adaptive solutions.

Tying all of this back to unconscious bias, there is little reason to assume that early humans developed specialized psychological modules to address social demands across the aforementioned dimensions. Broad selection pressures provide limited insight into the specific shape these bias mechanisms might have taken. Furthermore, specific selection pressures cannot be definitively known; even if they were, it would not guarantee that the inferred specialized mechanisms for categorizing individuals or generating prejudiced reactions indeed evolved. Finally, the environmental challenges influencing cognitive evolution were far from stable, suggesting that natural selection would have favored a flexible cognitive apparatus over the development of specialized modules that might become maladaptive in shifting

environments. Buller concludes that such evolutionary reasoning amounts to *"pure guesswork"*, thereby offering little confidence in the validity of inferred adaptive solutions. [86]. While this does not entirely discount the possibility that these hypotheses are valid, it imposes a highly rigorous burden of proof on the supporting empirical evidence [86].

7.4.3. Adaptive solutions

Evolutionary accounts propose that, through natural selection, humans developed specialized cognitive circuits (modules) designed to solve distinct problems within specific environmental domains [86]. Accordingly, ancestral populations became equipped with a multitude of modules, each responsible for domain-specific tasks, such as categorizing individuals by age, sex, group membership, and pathogen risk. These modules develop automatically, independently of exposure to their relevant domains, and possess intrinsic information about those specific environments. Moreover, modules operate relatively independently from other cognitive processes, which accounts for their capacity to process information rapidly within their domains. The logic underlying this modularity hypothesis rests on three broad arguments [109].

The first argument posits that just as natural selection has refined physical organs for highly specific functions, it must have also developed "mental organs" equipped to perform specialized tasks within distinct domains. Because general-purpose physical organs would be poorly suited to resolve diverse physiological challenges, general-purpose cognitive mechanisms would be equally ineffective at solving specific cognitive and behavioral tasks. Consequently, the human mind comprises a multitude of modules, each dedicated to a specific function. There are several reasons why this argument by analogy is flawed, but perhaps the most vivid counterexample is the immune system [86]. Successfully eliminating vastly diverse pathogens requires "domain-specific" responses from particular antibodies. However, these antibodies are produced by a "domain-general" immune system, which is not inherently specialized for any particular pathogen. As such, a general-purpose system can resolve highly specific problems by producing targeted, domain-specific responses. Accordingly, the human brain functions as a highly plastic, domain-general system capable of constructing "domain-dominant" cognitive circuits to address specific problem domains [86].

The second argument asserts that minds must solve diverse kinds of problems, each of them requiring specific solutions [86]. A general-purpose system would be ill-equipped to resolve such a vast array of problems, as no single problem-solving principle is universally applicable. Therefore, successfully addressing these unique challenges requires the use of specialized mechanisms, or modules, tailored to

these problem domains. However, similar to the first argument, a dedicated module is not required to produce a domain-specific solution, as a general-purpose processing system is equally capable of generating such responses. To clarify, there is a crucial difference between domain-specific processing and domain-specific knowledge. A domain-general mechanism of social learning can facilitate the acquisition of specialized knowledge, which subsequently generates optimal responses within that domain without requiring domain-specific processing [109].

The third argument claims that there is neither enough time nor information for a general-purpose mechanism to learn effective solutions across all relevant domains. Therefore, modules possessing innate knowledge are indispensable, as they provide immediate adaptive advantages by efficiently guiding behavior. The problem with this argument is two-fold [86]. First, a substantial amount of relevant knowledge can be socially acquired, whether transmitted by forebears or learned through direct observation. If this type of knowledge is sufficient to generate adaptive behavior, the necessity for a multitude of specialized modules is undermined. Second, general-purpose mechanisms come equipped with features that provide initial guidance regarding adaptively relevant domains. These *initial biases* nudge the developing brain to prioritize specific stimuli, eventually leading to the formation of specialized cognitive circuits that address particular adaptive challenges. For example, initial biases facilitate the development of a face recognition mechanism. Crucially, these innate predispositions do not dictate the final form of the circuit; instead, they lower the threshold for the brain to recognize and respond to relevant stimuli. In the absence of such stimuli, the corresponding neural circuits fail to develop.

Taking all of the above into consideration, it is highly improbable that the human brain has evolved specific, genetically heritable modules for categorizing individuals or generating affective responses. Instead, humans developed a plastic architecture capable of adapting to diverse environments [86]. Natural selection prioritized a highly flexible cognitive apparatus, one that fine-tunes its neural structure through the mechanisms of synaptic proliferation and pruning. Ultimately, environmental stimuli dictate the precise shape neural circuits take, partly guided by innate developmental predispositions. From this perspective, unconscious biases emerge not as inherited psychological modules, but as the consequence of domain-general learning processes that register the statistical regularities of their sociocultural environment. At most, inherited predispositions may coax the brain to attend to certain environmental cues, which can facilitate the development of particular biases; however, these predispositions lack the domain-specificity of evolutionary modules and do not determine the final form of a neural circuit. As such, the environment serves as the final arbiter of the brain's configuration.

7.4.4. Modern empirical evidence

The final pillar of evolutionary functional analysis involves conducting psychological experiments to validate hypotheses regarding psychological mechanisms that were inferred from ancestral adaptive challenges. As demonstrated throughout this section, the foundational premises of these hypotheses, along with the hypotheses themselves, are inherently problematic. As such, any empirical evidence cited in support of these hypotheses must meet a significantly higher standard of verification. Evolutionary psychologists claim that their empirical evidence is robust, citing consistent results across diverse cultures and populations. However, a fundamental question remains: do these experiments actually confirm the existence of the purported proximate mechanisms, or do they only reveal behavioral regularities that are not dependent on those specific mechanisms? To bypass this issue, the current subsection instead reviews findings from social neuroscience that directly assess the cognitive footprint of unconscious bias through neuroimaging and electrophysiological techniques. A review paper by neuroscientists Amodio and Cikara [110] synthesized the current neuroscientific evidence regarding social categorization, stereotypes, prejudice, and discrimination. Their findings offer no support for the existence of domain-specific mechanisms driving unconscious bias; instead, social cognition appears to be mediated by multiple general-purpose processes, which are summarized here:

1. Group-based categorizations are mediated by domain-general neural circuits responsible for mapping social regularities and assessing the functional significance of stimuli (functions primarily facilitated by the dorsal anterior cingulate cortex and anterior insula).
2. The brain relies on domain-general semantic memory (a conceptual storage network located in the anterior temporal lobe) to store and retrieve stereotypes and evaluations. Thus, the same cognitive system responsible for storing general knowledge also holds information regarding social groups and their associations with specific traits or affective values.
3. Prejudicial impressions are thought to be acquired through instrumental learning, a domain-general learning process (operated by the striatum). Through repeated interactions, specific actions directed toward social targets become associated with positive or negative reward outcomes, thus driving the formation of biased evaluations. This learning can occur without conscious awareness, and these impressions can be updated depending on the changing reward outcomes.
4. While the *amygdala* was previously regarded as a primary structure for implicit prejudice due to its role in reflexive fear processing (Table 5. Contrasting views on the *amygdala*), recent studies

have failed to conclusively support this view. Instead, *amygdala* activation within intergroup contexts is explained by its processing of behaviorally relevant stimuli, which shapes attitudes depending on whether the stimuli are advantageous or not. Thus, the *amygdala* is a domain-general relevance detector.

5. Discriminatory behavior is facilitated by a network of brain regions that translate prejudicial attitudes into social action. Prejudice provides the cognitive content that shapes how social information is processed across multiple domains. For example, it distorts face perception, leading to "perceptual dehumanization" of outgroup members [111], suppresses empathetic responses and heightens *Schadenfreude* toward outgroups, and biases decision-making to favor the ingroup while disadvantaging the outgroup.

Table 5. Contrasting views on the *amygdala*

Earlier view	**Current evidence**
Damage to *amygdala* impairs recognition of fear [112], hence *amygdala* is necessary for emotion of fear	Damage to the *amygdala* results in variable functional deficits: some individuals are capable of experiencing fear under specific circumstances [113], others become hypervigilant to threat [114], and some retain the ability to fully recognize fearful faces [115, 116]
The *amygdala* activates primarily to threatening stimuli	The *amygdala* also strongly reacts to images of food, sexual partners, and happy faces, provided they are intense or relevant to the subject's current goals [117]

7.4.5. Concluding the critique

As established throughout this section, evolutionary accounts of unconscious bias based on evolutionary functional analysis warrant serious skepticism. Broad characterizations of adaptive problems faced by ancestral populations are too imprecise to be informative, and there are no reliable methods for specifying them. Furthermore, the hypothesis that humans evolved specialized brain modules for specific problem domains is highly contested. Finally, contemporary neuroscientific evidence provides robust support for the domain-general nature of the mechanisms underlying unconscious bias, effectively refuting the modularity hypothesis.

An alternative interpretation proposes that natural selection favored a general-purpose, plastic brain architecture because ancestral populations encountered relatively volatile selection pressures concerning psychological evolution. This plasticity facilitated flexible phenotypes that proved more adaptive across shifting environmental conditions than rigid, modularized adaptations. In this view, the widespread phenomenon of unconscious bias represents a case of *phenotypic evolution*, reflecting how populations adapt to unpredictable and biased environments. To elaborate on this perspective, the final section of this review introduces a domain-general account of unconscious bias based on Predictive Processing, a highly influential framework in cognitive neuroscience [118].

7.5. Predictive Processing framework

7.5.1. Introduction to Predictive Processing

The primary function of the brain is to regulate the organism's physiological systems efficiently to ensure its chances at survival and reproduction [119]. The brain accomplishes this task best by anticipating the upcoming needs of its internal milieu and satisfying them before they arise. This proactive regulation is called allostasis [6]. To successfully anticipate these needs, the brain generates a comprehensive internal model of both the body and the external world [121]. This model forms the foundation of the Predictive Processing framework [13,122], which describes the brain as a probabilistic system that constantly makes predictions to minimize *surprise* – unexpected divergence from anticipated physiological needs.

The primary premise of this framework is that rather than passively reacting to bottom-up sensory input, the brain actively anticipates it by generating predictions. Operating on Bayesian principles, the brain estimates the probability of internal predictions given the incoming sensory data. It achieves this by continuously sending top-down prediction signals to lower sensory regions, registering any discrepant bottom-up input as a *prediction error*. This error represents the *surprise* that the brain constantly seeks to minimize. To resolve the error, the brain has two non-exclusive strategies: update its internal model to reflect sensory reality or initiate actions to change incoming sensory data so that it matches initial predictions [123]. The former process, which accounts for the sensory input that the model failed to

[6] The well-known Pavlov's dog experiment [120] is essentially an example of such anticipatory regulation, where the dog's digestive system is prepared for the anticipated food intake. Allostasis reconceptualizes the traditional homeostasis model, which relies on monitoring physiological parameters and correcting deviations from established set-points through reactive feedback mechanisms.

predict, is termed *perceptual inference*. The latter, through which incoming sensory data is actively shaped to match the prior prediction, is known as *active inference*.

7.5.2. The mechanism of stereotypes

According to Devine [124], people acquire beliefs about social groups from their surrounding culture. Due to frequent exposure to these associations, everyone, including non-prejudiced individuals, is familiarized with stereotypes, which are firmly lodged in semantic memory. Because individuals are often unaware of these latent stereotypes, they are classified as implicit biases.

The predictive processing framework explains this phenomenon in the following way [125]. The brain continuously updates its generative models to align with incoming sensory input. Through repeated exposure to pervasive cultural tropes and social associations, the brain infers statistical regularities linking specific social groups to particular traits and behaviors. These learned associations are encoded as prior predictions (*priors*) that guide subsequent perception and behavior toward these groups. Because predictive processes are mostly non-conscious [126], people are largely unaware of their internalized associations. Consequently, these priors can drive inadvertent discriminatory behavior, even among those who consciously disavow any personal prejudice [38].

Therefore, while the standard view within the social cognition perspective holds that stereotypes are manifestations of a cognitive bias [127], the predictive processing hypothesis contends that stereotypes are the result of the brain learning in environments that are inherently biased [125]. Given this view, stereotypes can be considered roughly accurate representations of the sociocultural environment [123]. For if stereotypes were inaccurate, the predictive brain would fail to reliably anticipate the world, leading to frequent prediction errors. Since the brain models the world to serve the body's physiological needs [102], and because prediction errors impose high metabolic costs on allostasis, these models are biologically pressured to remain accurate.

However, if stereotypes prove factually inaccurate, the brain relies on an alternative strategy to minimize prediction error. This is explained through the phenomenon of self-fulfilling prophecy, which acts through active inference. When encountering evidence that contradicts a stereotype, individuals may unconsciously adjust their behavior to provoke a response that confirms their initial prediction [7]. This interaction reinforces the original stereotype, increasing the likelihood that the same dynamic will repeat

[7] Perhaps updating the initial belief incurs higher metabolic costs than simply acting in a way that confirms the belief. This action can also entail not attending to disconfirming evidence in a broad sense.

in the future. Thus, even when faced with counter-stereotypical information, the brain can minimize prediction error and avoid high metabolic costs without updating its internal model, relying instead on actively shaping the social environment to align with its predictions [123].

To summarize, individuals acquire biased notions from their sociocultural environment. Through perceptual inference, the brain integrates these associations into its generative model as priors. In subsequent social encounters, the brain relies on these priors to minimize surprise, thereby reducing the metabolic costs of allostatic regulation. When a member from a social group exhibits traits that contradict these initial predictions, the brain may employ active inference to provoke behaviors consistent with the stereotype, thus resolving the prediction error. The following subsection further details how these predictive processes give rise to prejudice.

7.5.3. The mechanism of prejudice

To decide whether to utilize active or perceptual inference when encountering prediction error, the brain weighs the strength of its prediction against the reliability of the sensory evidence [123]. The higher the confidence in a prior prediction, the more likely the brain will rely on active inference instead of integrating the error to update the model. In this way, when individuals have strong initial beliefs about a social group, they may fail to entertain alternative hypotheses that might fit the sensory data better, thereby acting in a way that results in prejudice [128].

Simultaneously, the brain evaluates the reliability of the incoming sensory input, assigning *precision* to the resulting prediction error. The lower the precision assigned to an error signal, the more likely the brain is to dismiss incoming sensory input, relying on active inference to ensure the existing prediction dominates perception [128]. For example, when an individual is in a familiar environment, such as their own home, the brain may filter out incoming sensory data, instead relying on initial predictions to navigate the space. This same dynamic can lead to prejudice toward social groups perceived as "well-known". Conversely, when an individual encounters a novel situation, such as a back alley at night in a foreign city, the brain may assign high precision to the incoming error signals, resulting in heightened attention to sensory stimuli. In fact, attention and the precision assigned to an error signal are inextricably connected – as attention increases, so does precision [129]. As such, when people have little interest in a specific domain, such as marginalized populations, they may not pay attention to evidence that disconfirms their crude, prejudiced attitudes toward these groups. Therefore, prejudice emerges when strong initial beliefs about social groups are coupled with a failure to attend to disconfirming evidence [128].

Considered as a whole, the predictive processing perspective on unconscious bias suggests that perceptual inference leads to stereotypes and prejudice because the brain is learning in biased environments. Through active inference, these acquired biases are reinforced and perpetuated as individuals provoke stereotype-consistent behavior and overlook disconfirming evidence. Ultimately, unconscious bias functions as a strategy to minimize prediction error, thereby reducing the metabolic demands of allostatic regulation.

7.6. Discussion

Throughout this work, two conflicting interpretations of unconscious bias have emerged. One perspective, broadly speaking, conceptualizes bias as "given" – an inevitable feature of human cognition – whereas the other conceives of bias as "constructed" – inherently a product of sociocultural contingencies. The first view is reflected in theories that portray bias as efficient heuristics or innate, domain-specific evolutionary modules. The second view is apparent in theories that consider bias as a product of systemic inequality or as the result of the brain adapting to biased environments. The predictive processing framework, with its imperative for allostasis, bridges these two broad conceptions into a coherent whole: the brain's biological requirement for efficiency results in it internalizing regularities from environments pervaded by biases. Given that unconscious bias is a ubiquitous phenomenon with significant consequences for societal well-being, it is critical to identify the domains where interventions are most likely to yield meaningful change. Ultimately, the environment emerges as the central locus of intervention, as it determines the content of unconscious bias, the intensity of its prevalence, and the frequency of discriminatory behavior. Environments built on exploitation and intense competition will naturally generate a distinct bias profile compared to those based on cooperation and mutual understanding. Because humans have acquired an unprecedented capacity to shape the environments they inhabit, there remains a significant potential to reduce the prevalence and severity of unconscious bias through structural change. This potential is vividly encapsulated by Payne and colleagues [41]: "Far from being rigid [personal] attitudes, implicit biases could be like the weather: If you want it to change, simply wait a day or two. Average environmental levels of bias, in contrast, may be more like climate. Changes are slow, but even small changes may have immense consequences."

8. CONCLUSIONS

1) Unconscious bias is traditionally conceptualized as automatic, unintentional associations operating outside of conscious awareness. It encompasses three interconnected psychological processes: cognitive stereotypes, affective prejudices, and discriminatory behaviors. The literature on this topic is broad and often inconsistent. The terms "unconscious" and "implicit," though frequently used interchangeably, are technically distinct. The strict dichotomy between implicit and explicit cognition is increasingly contested by research demonstrating the fluid boundaries between the two. Some interpretations classify unconscious bias as part of cognitive biases, which function as fast, automatic heuristics, while others propose that these biases originate from the cultural landscape, functioning as individual-level indicators of systemic societal prejudice.

2) Unconscious bias poses a serious threat to public health by inflicting chronic stress and allostatic load on marginalized populations, while simultaneously driving cognitive heuristics that frequently lead to preventable medical errors in clinical settings.

3) Although evolutionary psychology has posited that unconscious biases are innate, domain-specific adaptations that arose in ancestral environments, philosophers of science rigorously question the validity of these underlying premises, and social neuroscience increasingly supports domain-general learning mechanisms instead. Ultimately, viewed through a predictive processing framework, unconscious bias emerges as a direct byproduct of the brain's biological imperative for allostatic efficiency, arising through the internalization of the statistical regularities of prejudiced sociocultural environments.

9. REFERENCES

1. Bristow W. Enlightenment. In, eds. Zalta EN, Nodelman U. *The Stanford Encyclopedia of Philosophy*. Fall 2023. Metaphysics Research Lab, Stanford University; 2023.

2. Plant EA, Devine PG. Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*. 1998;75(3):811-832. doi:10.1037/0022-3514.75.3.811

3. Fiske S. Stereotyping, prejudice, and discrimination. *The handbook of social psychology*. 1998;1.

4. Nosek BA, Smyth FL, Hansen JJ, et al. Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*. 2007;18:36-88.

5. Hewstone M, Rubin M, Willis H. Intergroup Bias. *Annual Review of Psychology*. 2002;53(1):575-604. doi:10.1146/annurev.psych.53.100901.135109

6. Devine PG, Plant EA, Amodio DM, Harmon-Jones E, Vance SL. The regulation of explicit and implicit race bias: The role of motivations to respond without prejudice. *Journal of Personality and Social Psychology*. 2002;82(5):835-848. doi:10.1037/0022-3514.82.5.835

7. Chekroud AM, Everett JAC, Bridge H, Hewstone M. A review of neuroimaging studies of race-related prejudice: Does amygdala response reflect threat? *Frontiers in Human Neuroscience*. 2014;8. doi:10.3389/fnhum.2014.00179

8. Eberhardt JL. Imaging race. *American Psychologist*. 2005;60(2):181-190. doi:10.1037/0003-066X.60.2.181

9. Greenwald AG, Dasgupta N, Dovidio JF, Kang J, Moss-Racusin CA, Teachman BA. Implicit-Bias Remedies: Treating Discriminatory Bias as a Public-Health Problem. *Psychological Science in the Public Interest*. 2022;23(1):7-40. doi:10.1177/15291006211070781

10. Forscher PS, Lai CK, Axt JR, et al. A meta-analysis of procedures to change implicit measures. *Journal of Personality and Social Psychology*. 2019;117(3):522-559. doi:10.1037/pspa0000160

11. Burke E. *Reflections on the Revolution in France*. (, ed.Turner FM). Yale University Press; 2008. doi:10.12987/9780300134865

12. Malpas J. Hans-georg gadamer. In, eds.Zalta EN, Nodelman U. *The Stanford Encyclopedia of Philosophy*. Fall 2025. Metaphysics Research Lab, Stanford University; 2025.

13. Clark A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*. 2013;36(3):181-204. doi:10.1017/S0140525X12000477

14. Freeman JB, Johnson KL. More Than Meets the Eye: Split-Second Social Perception. *Trends in Cognitive Sciences*. 2016;20(5):362-374. doi:10.1016/j.tics.2016.03.003

15. Heider F. *The Psychology of Interpersonal Relations.* John Wiley & Sons, Inc.; 1958. doi:10.1037/10628-000

16. Solms M. The scientific standing of psychoanalysis. *BJPsych international*. 2018;15(1):5-8. doi:10.1192/bji.2017.4

17. Bargh JA, Chartrand TL. The unbearable automaticity of being. *American Psychologist*. 1999;54(7):462-479. doi:10.1037/0003-066X.54.7.462

18. Lakoff G, Johnson M. *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. Basic Books; 1999.

19. Brownstein M. Implicit bias. In, eds.Zalta EN, Nodelman U. *The Stanford Encyclopedia of Philosophy*. Spring 2025. Metaphysics Research Lab, Stanford University; 2025.

20. Greenwald AG, Krieger LH. Implicit Bias: Scientific Foundations. *California Law Review*. 2006;94(4):945. doi:10.2307/20439056

21. Greenwald AG, Banaji MR. The implicit revolution: Reconceiving the relation between conscious and unconscious. *American Psychologist*. 2017;72(9):861-871. doi:10.1037/amp0000238

22. Goedderz A, Rahmani Azad Z, Hahn A. Awareness of Implicit Attitudes Revisited: A Meta-Analysis on Replications Across Samples and Settings, ed.Farb N. *Collabra: Psychology*. 2024;10(1):126220. doi:10.1525/collabra.126220

23. Nier JA. How Dissociated Are Implicit and Explicit Racial Attitudes? A Bogus Pipeline Approach. *Group Processes & Intergroup Relations*. 2005;8(1):39-52. doi:10.1177/1368430205048615

24. Hahn A, Judd CM, Hirsh HK, Blair IV. Awareness of implicit attitudes. *Journal of Experimental Psychology: General*. 2014;143(3):1369-1392. doi:10.1037/a0035028

25. Kurdi B, Melnikoff DE, Hannay JW, et al. Testing the automaticity features of the affect misattribution procedure: The roles of awareness and intentionality. *Behavior Research Methods*. 2023;56(4):3161-3194. doi:10.3758/s13428-023-02291-2

26. Greenwald AG, Lai CK. Implicit Social Cognition. *Annual Review of Psychology*. 2020;71(1):419-445. doi:10.1146/annurev-psych-010419-050837

27. Gawronski B, Brannon SM. Attitudes and the Implicit-Explicit Dualism. In: *The Handbook of Attitudes, Volume 1: Basic Principles*. 2nd ed. Routledge; 2018.

28. Gawronski B. Six Lessons for a Cogent Science of Implicit Bias and Its Criticism. *Perspectives on Psychological Science*. 2019;14(4):574-595. doi:10.1177/1745691619826015

29. Lai CK, Skinner AL, Cooley E, et al. Reducing implicit racial preferences: II. Intervention effectiveness across time. *Journal of Experimental Psychology: General*. 2016;145(8):1001-1016. doi:10.1037/xge0000179

30. Gawronski B, Morrison M, Phills CE, Galdi S. Temporal Stability of Implicit and Explicit Measures: A Longitudinal Analysis. *Personality and Social Psychology Bulletin*. 2017;43(3):300-312. doi:10.1177/0146167216684131

31. Sritharan R, Gawronski B. Changing Implicit and Explicit Prejudice: Insights from the Associative-Propositional Evaluation Model. *Social Psychology*. 2010;41(3):113-123. doi:10.1027/1864-9335/a000017

32. Payne BK. Implicit Bias: Evolution of a Powerful Idea. *Annual Review of Psychology*. 2026;77(1):593-617. doi:10.1146/annurev-psych-030525-043416

33. Fiske ST, Taylor SE. *Social Cognition: From Brains to Culture*. 3rd. edition. SAGE; 2017.

34. Dovidio JF, Hewstone M, Glick P, Esses VM. Prejudice, Stereotyping and Discrimination: Theoretical and Empirical Overview. In: *The SAGE Handbook of Prejudice, Stereotyping and Discrimination*. SAGE Publications Ltd; 2010:3-28. doi:10.4135/9781446200919.n1

35. Nadurak V. Heuristics and cognitive biases: A conceptual analysis. *Memory & Cognition*. Published online October 2025. doi:10.3758/s13421-025-01814-w

36. Gigerenzer G, Gaissmaier W. Heuristic Decision Making. *Annual Review of Psychology*. 2011;62(1):451-482. doi:10.1146/annurev-psych-120709-145346

37. Berthet V. The Impact of Cognitive Biases on Professionals' Decision-Making: A Review of Four Occupational Areas. *Frontiers in Psychology*. 2021;12:802439. doi:10.3389/fpsyg.2021.802439

38. Lausi G. How cognitive processes shape implicit stereotypes: A literature review. *Open Research Europe*. 2025;4:263. doi:10.12688/openreseurope.18691.2

39. Kahneman D. *Thinking, Fast and Slow*. Farrar, Straus and Giroux; 2011.

40. Bonilla-Silva E. Racial attitudes or racial ideology? An alternative paradigm for examining actors' racial views. *Journal of Political Ideologies*. 2003;8(1):63-82. doi:10.1080/13569310306082

41. Payne BK, Vuletich HA, Lundberg KB. The Bias of Crowds: How Implicit Bias Bridges Personal and Systemic Prejudice. *Psychological Inquiry*. 2017;28(4):233-248. doi:10.1080/1047840X.2017.1335568

42. Miller HN, LaFave S, Marineau L, Stephens J, Thorpe RJ. The impact of discrimination on allostatic load in adults: An integrative review of literature. *Journal of Psychosomatic Research*. 2021;146:110434. doi:10.1016/j.jpsychores.2021.110434

43. McEwen BS, Stellar E. Stress and the individual. Mechanisms leading to disease. *Archives of Internal Medicine*. 1993;153(18):2093-2101.

44. Guidi J, Lucente M, Sonino N, Fava GA. Allostatic Load and Its Impact on Health: A Systematic Review. *Psychotherapy and Psychosomatics*. 2021;90(1):11-27. doi:10.1159/000510696

45. Pascoe EA, Smart Richman L. Perceived discrimination and health: A meta-analytic review. *Psychological Bulletin*. 2009;135(4):531-554. doi:10.1037/a0016059

46. Inzlicht M, McKay L, Aronson J. Stigma as Ego Depletion: How Being the Target of Prejudice Affects Self-Control. *Psychological Science*. 2006;17(3):262-269. doi:10.1111/j.1467-9280.2006.01695.x

47. Govorun O, Payne BK. Ego – Depletion and Prejudice: Separating Automatic and Controlled Components. *Social Cognition*. 2006;24(2):111-136. doi:10.1521/soco.2006.24.2.111

48. Landrine H, Klonoff EA. The Schedule of Racist Events: A Measure of Racial Discrimination and a Study of Its Negative Physical and Mental Health Consequences. *Journal of Black Psychology*. 1996;22(2):144-168. doi:10.1177/00957984960222002

49. Clark R, Anderson NB, Clark VR, Williams DR. Racism as a stressor for African Americans: A biopsychosocial model. *American Psychologist*. 1999;54(10):805-816. doi:10.1037/0003-066X.54.10.805

50. FitzGerald C, Hurst S. Implicit bias in healthcare professionals: A systematic review. *BMC medical ethics*. 2017;18(1):19. doi:10.1186/s12910-017-0179-8

51. Croskerry P. Achieving Quality in Clinical Decision Making: Cognitive Strategies and Detection of Bias. *Academic Emergency Medicine*. 2002;9(11):1184-1204. doi:10.1197/aemj.9.11.1184

52. O'Sullivan E, Schofield S. Cognitive Bias in Clinical Medicine. *Journal of the Royal College of Physicians of Edinburgh*. 2018;48(3):225-232. doi:10.4997/jrcpe.2018.306

53. Oxtoby K. How unconscious bias can discriminate against patients and affect their care. *BMJ*. Published online November 2020:m4152. doi:10.1136/bmj.m4152

54. Shah HS, Bohlen J. Implicit Bias. In: *StatPearls*. StatPearls Publishing; 2026.

55. Kaur Dhillon J. Psychological Effect of Medical Gaslighting on Female Patients: A Systematic Review. *International Journal of Innovative Science and Research Technology*. Published online June 2025:845-852. doi:10.38124/ijisrt/25jun646

56. Biddle C, Fallavollita JA, Homish GG, Orom H. Gender bias in clinical decision making emerges when patients with coronary heart disease symptoms also have psychological symptoms. *Heart & Lung*. 2019;48(4):331-338. doi:10.1016/j.hrtlng.2018.11.005

57. Johnson HM, Gorre CE, Friedrich-Karnik A, Gulati M. Addressing the Bias in Cardiovascular Care: Missed & Delayed Diagnosis of Cardiovascular Disease in Women. *American Journal of Preventive Cardiology*. 2021;8:100299. doi:10.1016/j.ajpc.2021.100299

58. Liu KA, Mager NAD. Women's involvement in clinical trials: Historical perspective and future implications. *Pharmacy Practice*. 2016;14(1):708. doi:10.18549/PharmPract.2016.01.708

59. Wallis CJD, Jerath A, Coburn N, et al. Association of Surgeon-Patient Sex Concordance With Postoperative Outcomes. *JAMA Surgery*. 2022;157(2):146. doi:10.1001/jamasurg.2021.6339

60. Vargas AJ, Tobey-Moore L, Curran GM, et al. Exploring Racial Disparities in Chronic Pain Management. *Journal of Pain Research*. 2025;18:2901-2908. doi:10.2147/JPR.S493664

61. Schulman KA, Berlin JA, Harless W, et al. The Effect of Race and Sex on Physicians' Recommendations for Cardiac Catheterization. *New England Journal of Medicine*. 1999;340(8):618-626. doi:10.1056/NEJM199902253400806

62. Zavala VA, Bracci PM, Carethers JM, et al. Cancer health disparities in racial/ethnic minorities in the United States. *British Journal of Cancer*. 2021;124(2):315-332. doi:10.1038/s41416-020-01038-6

63. Blair IV, Steiner JF, Fairclough DL, et al. Clinicians' Implicit Ethnic/Racial Bias and Perceptions of Care Among Black and Latino Patients. *The Annals of Family Medicine*. 2013;11(1):43-52. doi:10.1370/afm.1442

64. Jones S, Howard L, Thornicroft G. "Diagnostic overshadowing": Worse physical health care for people with mental illness. *Acta Psychiatrica Scandinavica*. 2008;118(3):169-171. doi:10.1111/j.1600-0447.2008.01211.x

65. Freeman J, Strauss P, Hamilton S, et al. They Told Me "This Isn't a Hotel": Young People's Experiences and Perceptions of Care When Presenting to the Emergency Department with Suicide-Related Behaviour. *International Journal of Environmental Research and Public Health*. 2022;19(3):1377. doi:10.3390/ijerph19031377

66. Chang ES, Kannoth S, Levy S, Wang SY, Lee JE, Levy BR. Global reach of ageism on older persons' health: A systematic review, ed.Bayer A. *PLOS ONE*. 2020;15(1):e0220857. doi:10.1371/journal.pone.0220857

67. Valdiserri RO. HIV/AIDS Stigma: An Impediment to Public Health. *American Journal of Public Health*. 2002;92(3):341-342. doi:10.2105/AJPH.92.3.341

68. Katz IT, Ryu AE, Onuegbu AG, et al. Impact of HIV-related stigma on treatment adherence: Systematic review and meta-synthesis. *Journal of the International AIDS Society*. 2013;16(3S2):18640. doi:10.7448/IAS.16.3.18640

69. Casanova-Perez R, Apodaca C, Bascom E, et al. Broken down by bias: Healthcare biases experienced by BIPOC and LGBTQ+ patients. *AMIA Annual Symposium proceedings AMIA Symposium*. 2021;2021:275-284.

70. Van Boekel LC, Brouwers EPM, Van Weeghel J, Garretsen HFL. Stigma among health professionals towards patients with substance use disorders and its consequences for healthcare delivery: Systematic review. *Drug and Alcohol Dependence*. 2013;131(1-2):23-35. doi:10.1016/j.drugalcdep.2013.02.018

71. Implicit preadmission screening criteria in an alcoholism treatment program. *American Journal of Psychiatry*. 1979;136(9):1190-1193. doi:10.1176/ajp.136.9.1190

72. Pettit KE, Turner JS, Kindrat JK, et al. Effect of Socioeconomic Status Bias on Medical Student-Patient Interactions Using an Emergency Medicine Simulation. *AEM education and training*. 2017;1(2):126-131. doi:10.1002/aet2.10022

73. Gilmer C, Buccieri K. Homeless Patients Associate Clinician Bias With Suboptimal Care for Mental Illness, Addictions, and Chronic Pain. *Journal of Primary Care & Community Health*. 2020;11:2150132720910289. doi:10.1177/2150132720910289

74. Reichard A, Stolzle H, Fox MH. Health disparities among adults with physical disabilities or cognitive limitations compared to individuals with no disabilities in the United States. *Disability and Health Journal*. 2011;4(2):59-67. doi:10.1016/j.dhjo.2010.05.003

75. American Association on Intellectual and Developmental Disabilities, American Autoimmune Related Diseases Association, American Foundation for the Blind, et al. Eliminating Disability Bias. *Health Affairs*. 2021;40(5):851-851. doi:10.1377/hlthaff.2021.00409

76. Kirschner KL. Educating Health Care Professionals to Care for Patients With Disabilities. *JAMA*. 2009;302(12):1334. doi:10.1001/jama.2009.1398

77. Murphy J, Farrell K, Kealy MB, Kristiniak S. Mindfulness as a self-care strategy for healthcare professionals to reduce stress and implicit bias. *Journal of Interprofessional Education & Practice*. 2023;30:100598. doi:10.1016/j.xjep.2022.100598

78. Goulart MF, Huayllani MT, Balch Samora J, Moore AM, Janis JE. Assessing the Prevalence of Microaggressions in Plastic Surgery Training: A National Survey. *Plastic and Reconstructive Surgery - Global Open*. 2021;9(12):e4062. doi:10.1097/GOX.0000000000004062

79. Gonzalez CM, Deno ML, Kintzer E, Marantz PR, Lypson ML, McKee MD. Patient perspectives on racial and ethnic implicit bias in clinical encounters: Implications for curriculum development. *Patient Education and Counseling*. 2018;101(9):1669-1675. doi:10.1016/j.pec.2018.05.016

80. Sng O, Williams KEG, Neuberg SL. Evolutionary Approaches to Stereotyping and Prejudice. In, eds.Sibley CG, Barlow FK. *The Cambridge Handbook of the Psychology of Prejudice*. 1st ed. Cambridge University Press; 2016:21-46. doi:10.1017/9781316161579.002

81. Neuberg SL, Schaller M. An evolutionary threat-management approach to prejudices. *Current Opinion in Psychology*. 2016;7:1-5. doi:10.1016/j.copsyc.2015.06.004

82. Cottrell CA, Park JH. Evolutionary perspectives on prejudice. In: *Stereotyping and Prejudice*. Frontiers of social psychology. Psychology Press; 2013:29-51.

83. Moore DS. *The Developing Genome: An Introduction to Behavioral Epigenetics.* Oxford University Press; 2015:x, 308.

84. Mayer GC, Craig CL. Evolution, Theory of. In: *Encyclopedia of Biodiversity*. Elsevier; 2013:392-400. doi:10.1016/B978-0-12-384719-5.00048-4

85. Müller GB. Why an extended evolutionary synthesis is necessary. *Interface Focus*. 2017;7(5):20170015. doi:10.1098/rsfs.2017.0015

86. Buller DJ. *Adapting Minds: Evolutionary Psychology and the Persistent Quest for Human Nature.* Boston Review; 2005:xi, 550.

87. Bentham J, Di Cesare M, Stevens G, et al. A century of trends in adult human height. *eLife Sciences*. 2016;5.

88. Tooby J, Cosmides L. On the Universality of Human Nature and the Uniqueness of the Individual: The Role of Genetics and Adaptation. *Journal of Personality*. 1990;58(1):17-67. doi:10.1111/j.1467-6494.1990.tb00907.x

89. Cohen KM, Finney SC, Gibbard PL, Fan J-X. The ICS International Chronostratigraphic Chart. *Episodes*. 2013;36(3):199-204. doi:10.18814/epiiugs/2013/v36i3/002

90. Brewer MB. The importance of being we: Human nature and intergroup relations. *American Psychologist*. 2007;62(8):728-738. doi:10.1037/0003-066X.62.8.728

91. Stearns SC. Life history evolution: Successes, limitations, and prospects. *Naturwissenschaften*. 2000;87(11):476-486. doi:10.1007/s001140050763

92. Williams KEG, Sng O, Neuberg SL. Ecology-driven stereotypes override race stereotypes. *Proceedings of the National Academy of Sciences*. 2016;113(2):310-315. doi:10.1073/pnas.1519401113

93. Kurzban R, Tooby J, Cosmides L. Can race be erased? Coalitional computation and social categorization. *Proceedings of the National Academy of Sciences*. 2001;98(26):15387-15392. doi:10.1073/pnas.251541498

94. Pietraszewski D, Cosmides L, Tooby J. The content of our cooperation, not the color of our skin: An alliance detection system regulates categorization by coalition and race, but not sex. *PloS One*. 2014;9(2):e88534. doi:10.1371/journal.pone.0088534

95. Griskevicius V, Tybur JM, Delton AW, Robertson TE. The influence of mortality and socioeconomic status on risk and delayed rewards: A life history theory approach. *Journal of Personality and Social Psychology*. 2011;100(6):1015-1026. doi:10.1037/a0022403

96. Massey D. Geographies of responsibility. *Geografiska Annaler: Series B, Human Geography*. 2004;86(1):5-18. doi:10.1111/j.0435-3684.2004.00150.x

97. Schaller M, Park JH. The Behavioral Immune System (and Why It Matters). *Current Directions in Psychological Science*. 2011;20(2):99-103. doi:10.1177/0963721411402596

98. Haselton MG, Nettle D. The Paranoid Optimist: An Integrative Evolutionary Model of Cognitive Biases. *Personality and Social Psychology Review*. 2006;10(1):47-66. doi:10.1207/s15327957pspr1001_3

99. Nesse RM. The smoke detector principle. Natural selection and the regulation of defensive responses. *Annals of the New York Academy of Sciences*. 2001;935:75-85.

100. McDonald MM, Navarrete CD, Van Vugt M. Evolution and the psychology of intergroup conflict: The male warrior hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2012;367(1589):670-679. doi:10.1098/rstb.2011.0301

101. Cosmides L, Tooby J. Origins of domain specificity: The evolution of functional organization. In, eds.Hirschfeld LA, Gelman SA. *Mapping the Mind*. 1st ed. Cambridge University Press; 1994:85-116. doi:10.1017/CBO9780511752902.005

102. Barrett LF. Functionalism cannot save the classical view of emotion. *Social Cognitive and Affective Neuroscience*. 2017;12(1):34-36. doi:10.1093/scan/nsw156

103. Kelemen D, Rottman J, Seston R. Professional physical scientists display tenacious teleological tendencies: Purpose-based reasoning as a cognitive default. *Journal of Experimental Psychology: General*. 2013;142(4):1074-1083. doi:10.1037/a0030399

104. Vallacher RR, Wegner DM. What do people think they're doing? Action identification and human behavior. *Psychological Review*. 1987;94(1):3-15. doi:10.1037/0033-295X.94.1.3

105. Sireikis A. *Struktūrinė Literatūros Psichoanalizė: Diskursyvinis Tapatinimasis*. Humanities. Vilnius University; 2025. doi:10.15388/vu.thesis.744

106. Mayr E. Cause and effect in biology. *Science*. 1961;134(3489):1501-1506. Accessed April 2, 2026. https://www.jstor.org/stable/1707986

107. Sterelny K, Griffiths PE, Griffiths P. *Sex and Death: An Introduction to Philosophy of Biology*. University of Chicago press; 1999.

108. Gould SJ, Lewontin RC. The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society of London Series B Biological Sciences*. 1979;205(1161):581-598. doi:10.1098/rspb.1979.0086

109. Downes SM. Evolutionary psychology. In, eds.Zalta EN, Nodelman U. *The Stanford Encyclopedia of Philosophy*. Spring 2024. Metaphysics Research Lab, Stanford University; 2024.

110. Amodio DM, Cikara M. The Social Neuroscience of Prejudice. Published online 2020.

111. Fincher KM, Tetlock PE. Perceptual dehumanization of faces is activated by norm violations and facilitates norm enforcement. *Journal of Experimental Psychology: General*. 2016;145(2):131-146. doi:10.1037/xge0000132

112. Feinstein JS, Adolphs R, Damasio A, Tranel D. The Human Amygdala and the Induction and Experience of Fear. *Current Biology*. 2011;21(1):34-38. doi:10.1016/j.cub.2010.11.042

113. Adolphs R, Gosselin F, Buchanan TW, Tranel D, Schyns P, Damasio AR. A mechanism for impaired fear recognition after amygdala damage. *Nature*. 2005;433(7021):68-72. doi:10.1038/nature03086

114. Terburg D, Morgan BE, Montoya ER, et al. Hypervigilance for fear after basolateral amygdala damage in humans. *Translational Psychiatry*. 2012;2(5):e115-e115. doi:10.1038/tp.2012.46

115. Pishnamazi M, Tafakhori A, Loloee S, et al. Attentional bias towards and away from fearful faces is modulated by developmental amygdala damage. *Cortex*. 2016;81:24-34. doi:10.1016/j.cortex.2016.04.012

116. Tsuchiya N, Moradi F, Felsen C, Yamazaki M, Adolphs R. Intact rapid detection of fearful faces in the absence of the amygdala. *Nature Neuroscience*. 2009;12(10):1224-1225. doi:10.1038/nn.2380

117. Cunningham WA, Brosch T. Motivational Salience: Amygdala Tuning From Traits, Needs, Values, and Goals. *Current Directions in Psychological Science*. 2012;21(1):54-59. doi:10.1177/0963721411430832

118. Walsh KS, McGovern DP, Clark A, O'Connell RG. Evaluating the neurophysiological evidence for predictive processing as a model of perception. *Annals of the New York Academy of Sciences*. 2020;1464(1):242-268. doi:10.1111/nyas.14321

119. Sterling P, Laughlin S. *Principles of Neural Design*. The MIT Press; 2015. doi:10.7551/mitpress/9780262028707.001.0001

120. Pavlov I. Physiology of digestion. Nobel Lecture. Published online 1904.

121. Barrett LF. The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*. Published online October 2016:nsw154. doi:10.1093/scan/nsw154

122. Friston K. The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*. 2010;11(2):127-138. doi:10.1038/nrn2787

123. Villiger D. Stereotypes and self-fulfilling prophecies in the Bayesian brain. *Inquiry*. 2025;68(10):3317-3341. doi:10.1080/0020174X.2023.2166983

124. Devine PG. Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*. 1989;56(1):5-18. doi:10.1037/0022-3514.56.1.5

125. Hinton P. Implicit stereotypes and the predictive brain: Cognition and culture in "biased" person perception. *Palgrave Communications*. 2017;3(1):17086. doi:10.1057/palcomms.2017.86

126. Clark A. *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford University Press USA; 2016.

127. Fiske ST, Taylor SE. *Social Cognition: From Brains to Culture*. SAGE Publications Ltd; 2013. doi:10.4135/9781529681451

128. Hung TW. Why Human Prejudice is so Persistent: A Predictive Coding Analysis. *Social Epistemology*. 2023;37(6):779-797. doi:10.1080/02691728.2023.2237942

129. Feldman H, Friston KJ. Attention, Uncertainty, and Free-Energy. *Frontiers in Human Neuroscience*. 2010;4. doi:10.3389/fnhum.2010.00215