Curiosity

Pierre-Yves Oudeyer
Inria and University of Bordeaux, France
http://www.pyoudeyer.com

Note: This document is a preprint, all comments and suggestions for improvement are welcome

Abstract: *Curiosity — the set of processes that drive organisms to spontaneously explore and learn about their environment — is a fundamental aspect of human and animal cognition. It is also a growing focus of research across psychology, neuroscience, artificial intelligence, and education. This review provides an interdisciplinary overview of what science currently knows about curiosity, how different fields conceptualize and study it, and why it matters for society.*

The concept is traced from its philosophical origins to modern experimental and computational approaches, and core theoretical frameworks that have shaped the field are presented. As a particular form of intrinsic motivation — the drive to engage in activities for their inherent satisfaction rather than for external rewards — curiosity plays a distinctive role in how organisms organize their own exploration. A central theme is that curiosity is best understood not as a single emotion or fixed personality trait, but as a family of diverse processes driving spontaneous exploration across very different timescales — from rapid attention shifts toward novel stimuli lasting a few seconds, to sustained intellectual pursuits unfolding over months or years. Computational models have played a uniquely productive role in formalizing and unifying these diverse curiosity processes, generating testable predictions, and bridging disciplines.

The review also examines how scientists understand what curiosity does for individuals during their lifetime — such as enabling efficient learning and the acquisition of broad and diverse skill repertoires — as well as why curiosity evolved and how it may contribute to innovation across generations. The neural mechanisms underlying curiosity, its development across the lifespan, and its expression in non-human animals are discussed. Current scientific debates are also addressed, including which intrinsic reward signals best explain human exploration, how curiosity interacts with metacognition, agency and social interaction, and why new experimental paradigms are needed to study open-ended, self-directed exploration.

Finally, the review highlights two domains where understanding curiosity carries broad societal implications. In education, a growing body of research reveals how curiosity enhances learning and memory, and points toward concrete opportunities for improving classroom practices to nurture curiosity and reduce educational inequalities. In artificial intelligence, curiosity-driven exploration algorithms have proven essential for enabling machines to explore autonomously and learn efficiently, opening perspectives toward building systems capable of open-ended learning and discovery. This review is intended for a wide readership,

including graduate students and researchers across the cognitive sciences, as well as educators and practitioners seeking a synthesis of this rapidly evolving field.

1. Definition(s)

Curiosity refers to a set of processes that push organisms to spontaneously explore their environment for the sake of acquiring knowledge and competence. Curiosity is a particular form of intrinsic motivation: it is a motivational system that gets individuals engaged in exploring situations or problems for the purpose of building models of the world, and/or building diverse repertoires of skills, as opposed to other forms of exploration driven by extrinsic objectives such as obtaining food or social recognition.

The set of processes associated with curiosity is diverse, ranging from mechanisms that orient individual's attention and exploration towards novel or surprising stimuli within time spans of a few seconds, to searching information to fill a knowledge gap over a few minutes, and to mechanisms enabling individuals to self-generate and select their own problems/goals, engaging over longer time spans for the pleasure of making progress in learning.

The term "curiosity" has also been used to refer to the set of emotional states and feelings associated with these spontaneous exploration processes, as well as to general personality traits associated with individuals propensity to spontaneously explore and their general preferences over certain forms of curiosity-driven exploration.

In general, curiosity processes strongly interact with learning processes, and with metacognitive processes. They can also be influenced by, and influence, interaction with social peers.

Historically, the term curiosity has first been used in everyday language, to speak informally about human behaviour and its intuitive interpretation, i.e. it has often been used as a folk psychology term. This has led to difficulties in establishing and agreeing on precise definitions in the scientific community, as well as to overlap with related concepts such as "interest" and "intrinsic motivation". However, this has not been a major obstacle to making progress in measuring, modeling and theorizing the diversity of mechanisms involved in spontaneous exploration, across psychology, neuroscience and artificial intelligence modeling.

2. History

2.1 Early discussions in history

The concept of curiosity has a complex historical lineage. In Ancient Greece, while there was no direct equivalent to our modern term "curiosity," philosophers recognized forms of

knowledge-seeking with varied connotations. Aristotle's famous opening to his Metaphysics—"all men by nature desire to know"—acknowledged an innate drive toward knowledge acquisition. The Latin term "curiositas" emerged during the Roman period with multifaceted meanings: sometimes used positively for scholarly inquiry, sometimes as a distracting pursuit of unnecessary knowledge, or with negative connotations of inappropriate inquisitiveness or meddling (Harrison, 2001). This ambivalence toward curiosity persisted through the Middle Ages, with religious authorities frequently warning against excessive inquiry into divine matters (Ball, 2012; Daston & Park, 1998). The Renaissance and Enlightenment witnessed a gradual rehabilitation of curiosity as a driver of scientific inquiry (Ball, 2012).
In 1891, William James characterized curiosity as an instinct that evolved to facilitate survival and adaptation through active exploration of the environment. This evolutionary framing highlighted curiosity's adaptive value, suggesting it played a crucial role in helping organisms navigate and learn about their surroundings. Early educational philosophers like Maria Montessori further emphasized curiosity's importance, advocating for learning environments that nurture children's natural exploratory tendencies.

2.2 Foundational intuitions in psychology

Psychology's systematic study of curiosity emerged in the 1950s and 1960s, with multiple pioneers.

D.E. Berlyne (1960, 1965) established several foundational distinctions that continue to shape curiosity research today. He differentiated between epistemic curiosity (directed toward knowledge acquisition) and perceptual curiosity (directed toward sensory exploration), as well as between specific curiosity (targeted at resolving particular uncertainty) and diversive curiosity (seeking stimulation to alleviate boredom). Berlyne's framework was also particularly influential for introducing an information-theoretic approach to curiosity. His concept of "collative variables"—properties like novelty, surprise, complexity, ambiguity, and incongruity—emphasized that curiosity is driven by the comparative processing of information from different sources or time points. Importantly, these collative variables generate arousal and become stimuli for exploratory behavior when they create what Berlyne termed "conceptual conflict." This information-processing framework was remarkably prescient, anticipating contemporary computational theories that quantify curiosity in terms of uncertainty reduction, prediction errors, and information gain (Kidd and Hayden, 2015; Oudeyer, 2018; Dubey and Griffiths, 2020).

Early psychological investigations conceptualized curiosity variously. Some theorists approached curiosity through the drive theory framework (Hull, 1943), with Montgomery (1954) proposing a "drive for exploration" and Harlow (1950) a "drive to manipulate." However, White (1959) criticized this approach, noting that exploratory behaviors are not homeostatic like traditional drives - they don't diminish when satisfied but often intensify with engagement. White's alternative "competence motivation" framing positioned curiosity as intrinsically motivated behavior directed toward mastery, a view later elaborated by De Charms (1968) through concepts of personal causation and self-determination.

This emphasis on intrinsic motivation became foundational to Self-Determination Theory (Deci & Ryan, 1985, Deci et al., 2001), which distinguished intrinsic motivation (engaging in activities for their inherent satisfaction) from extrinsic motivation (driven by separable outcomes, such as food or money). SDT positioned curiosity as a prototypical form of intrinsic motivation, facilitated by environments supporting psychological needs for autonomy, competence, and relatedness. Experimentally, Deci (1971) demonstrated that external rewards could undermine intrinsic motivation for inherently interesting activities, revealing complex dynamics between curiosity and incentive systems—findings with significant implications for educational practice.

Alternative frameworks emerged, including Festinger's (1957) cognitive dissonance theory, which portrayed curiosity as motivated by a need to reduce inconsistencies between cognitive structures. Kagan (1972) and Loewenstein (1994) further developed this uncertainty-reduction view, with Loewenstein's "knowledge gap" theory characterizing curiosity as an aversive feeling of information deprivation. However, these theories inadequately explained why people sometimes voluntarily increase uncertainty through exploration.

Another influential perspective focused on optimal stimulation levels. Hunt (1965) suggested children seek "optimal incongruity," while Berlyne (1960) observed that people prefer intermediate levels of novelty. This "Goldilocks principle" was later empirically supported by findings that infants attend preferentially to stimuli of moderate complexity (Kidd et al., 2012). Similarly, theories of competence motivation (White, 1959; De Charms, 1968; Deci & Ryan, 1985) framed curiosity as a desire for mastery and control, resonating with Csikszentmihalyi's (1991) concept of "flow," the optimal experience achieved when challenge aligns with skill level.

2.3 Obstacles to the study of curiosity in the 20th century

Despite its fundamental importance in learning, the mechanisms of curiosity remained experimentally understudied throughout much of the 20th century. A significant obstacle was methodological: traditional experimental paradigms in psychology and neuroscience typically imposed predefined problems on participants, together with constraints on the space of options they could consider to solve these problems, fundamentally incompatible with studying curiosity as self-determined exploration and question-asking in an open world. Laboratory settings further constrained natural exploratory behavior, while behaviorist perspectives in psychology directed focus away from "mentalistic" constructs like curiosity and toward directly observable responses to stimuli. The emphasis on controlled laboratory conditions offering high internal validity came at the expense of ecological validity, limiting understanding of how curiosity manifests in naturalistic settings. These methodological constraints made it difficult to capture curiosity's most distinctive feature: its autonomous, self-directed nature.

2.4 History of research on non-human animal curiosity

Research on animal curiosity has roots in early comparative psychology, with Willard Small's 1899 study of white rats noting curiosity as their "most striking" intellectual trait. He observed that rats' curiosity developed early and often overrode fear. Subsequent research demonstrated that rats would actively explore mazes even when well-fed, and would temporarily ignore food to investigate novel environmental changes (Dashiell, 1925). Particularly revealing were studies showing that rats would cross electrified floors to explore novel areas (Warden, 1931), suggesting curiosity's motivational strength. This exploratory drive was not random: rats systematically favored novel over familiar locations (Dennis, 1934; Montgomery, 1952). Similar curiosity-driven behaviors have been documented across diverse species, from primates to birds and cephalopods. Hughes (2007) reported that rats consistently preferred obtaining novel stimuli over obtaining food or drugs, sometimes even crossing electrifying grids, further demonstrating curiosity's powerful motivational force across species.

2.5 The sciences of curiosity: an emerging domain to naturalize the study of curiosity

The turn of the 21st century marked a renaissance in curiosity research, with interdisciplinary approaches offering new methodological and conceptual tools. Around 2000, researchers began systematically investigating curiosity's neural, cognitive, and computational mechanisms, establishing a truly interdisciplinary science of curiosity.

Advances in neuroimaging techniques enabled researchers to examine the neural correlates of curiosity states in humans, while single-cell recording studies revealed how information-seeking behavior is processed in non-human primates. Developmental robotics research demonstrated how curiosity-like mechanisms could drive autonomous learning in artificial agents, offering computational insights into curiosity's functional architecture.

Computational modeling emerged as a particularly fruitful approach, providing formal frameworks to simulate and test theories of curiosity-driven learning. These models helped bridge gaps between disciplines by providing a common theoretical language and generating testable predictions about how curiosity influences learning across different contexts and species.

Recent work in machine learning and artificial intelligence has further emphasized curiosity's importance, with curiosity-inspired algorithms proving essential for problems requiring autonomous exploration in complex environments.

This interdisciplinary convergence of computational, neuroscientific, and psychological approaches has enabled more naturalistic studies of curiosity, moving beyond simplified laboratory paradigms to examine how curiosity operates in rich, ecological contexts.

3. Core Concepts of Curiosity

3.1 Curiosity and spontaneous exploration

Curiosity, as a general term often used in folk psychology —the everyday understanding and interpretation of mental states and behaviors—, encompasses diverse forms of spontaneous exploration that manifest across vastly different timescales and contexts. At its core, curiosity drives organisms to actively seek out and engage with their environment without immediate extrinsic rewards, ranging from reflexive orienting responses lasting mere seconds to sustained investigative behaviors spanning years. On short timescales, curiosity manifests as rapid attention shifts toward surprising stimuli—such as turning toward an unexpected noise or visually tracking a novel object entering one's field of view. These brief exploratory episodes can cascade into more extended forms of investigation, such as manipulating unfamiliar objects, systematically exploring new environments, or pursuing answers to specific questions over minutes or hours. At the longest timescales, curiosity underlies complex human endeavors like scientific research programs, artistic exploration, or lifelong learning pursuits, where individuals may dedicate years to understanding phenomena or mastering skills purely for the intrinsic satisfaction of discovery and understanding.

3.2 Curiosity as a special form of intrinsic motivation

Beyond folk psychological descriptions, curiosity can be scientifically approached through the conceptual framework of intrinsic motivation, which provides a foundation for understanding its mechanisms and functions. Intrinsic motivation refers to engagement in activities for their inherent satisfaction—the enjoyment, interest, or sense of competence they provide—rather than for separable consequences or external rewards (Deci & Ryan, 1985). This contrasts with extrinsic motivation, where behavior is driven by external outcomes such as obtaining rewards (money, food, praise) or avoiding punishments. Within the broader category of intrinsic motivation, which includes diverse phenomena like enjoyment of familiar activities or flow states during skilled performance (Csikszentmihalyi, 1991), curiosity represents a special subfamily specifically oriented toward exploration and learning (Oudeyer & Kaplan, 2007). What distinguishes curiosity from other forms of intrinsic motivation is its focus on seeking novelty, reducing uncertainty, and acquiring new knowledge or skills—it is the intrinsic motivation to explore the unknown rather than to engage with the familiar.

This conceptualization aligns with theoretical accounts that characterize curiosity as non-instrumental information seeking, where individuals pursue information purely for the satisfaction of acquiring knowledge, without any immediate practical application (Loewenstein, 1994; Kidd & Hayden, 2015). An example of non-instrumental information seeking (intrinsically motivated) is when someone reads about a topic simply because they find it fascinating, while instrumental information seeking (extrinsically motivated) happens when gathering information serves a specific external goal, such as researching restaurants to plan a dinner. However, this distinction is not always clear-cut in practice: many real-world exploratory behaviors are

simultaneously driven by both curiosity and extrinsic goals (Murayama et al., 2019; Oudeyer and Kaplan, 2007). A scientist might be genuinely curious about natural phenomena while also motivated by career advancement; a child exploring a new toy might be intrinsically interested in its properties while also seeking parent approval. This dual motivation is common in educational settings, where learning activities ideally engage students' natural curiosity while also serving instrumental goals like exam preparation (Deci et al., 2001). Understanding curiosity as a form of intrinsic motivation thus provides a framework for examining how internal drives for exploration interact with external incentives, and how environments can be designed to support both.

3.3 Curiosity as state and metacognitive feeling

Beyond viewing curiosity as a motivational process, researchers have conceptualized it as a metacognitive feeling—a subjective experience that provides information about one's own cognitive states and learning opportunities (Goupil & Proust, 2023). Metacognitive feelings, like confidence, familiarity, or tip-of-the-tongue experiences, serve as internal signals that guide cognitive behavior, sometimes without requiring explicit reasoning about mental states. In this framework, the feeling of curiosity emerges when individuals detect knowledge gaps, prediction errors, or potential learning opportunities, functioning as an affective marker that something is worth exploring or understanding better. This state-based conceptualization distinguishes momentary experiences of curiosity from stable personality traits, focusing on the fluctuating subjective sensations that arise in response to specific stimuli or situations.

Measuring curiosity as a metacognitive state requires capturing these transient experiences through various methodological approaches (Jirout et al., 2024). Self-report measures include experience sampling methods where participants rate their momentary curiosity levels throughout daily activities, or post-task questionnaires assessing felt curiosity during specific learning episodes (Ryan, 1982; Litman et al. 2005). Behavioral indicators of curiosity states include information-seeking actions (clicking for more information, Jirout and Evans, 2023; question-asking, Goupil and Proust, 2023)), exploration time allocation (Abdelghani et al. 2022) or exploration patterns (Lydon-Staley et al., 2021), and willingness to sacrifice rewards for information (Bennett et al., 2016; Embrey et al., 2024). Physiological measures such as pupil dilation and gaze patterns (Baranes et al., 2015), as well as neuroimagery have also been used to measure and characterize curiosity states (Kang et al., 2009; Gruber et al., 2014). Recent paradigms combine these approaches, using trivia questions or visual stimuli to elicit curiosity, then measuring both subjective ratings and subsequent memory performance to understand how metacognitive feelings of curiosity influence learning (Gruber et al., 2014; Fastrich et al., 2018). These multi-method approaches reveal that conceptualizing curiosity as a metacognitive feeling involves both the subjective experience of feeling curious ("I feel curious about this") and functional consequences for attention and memory, bridging subjective experience with objective learning outcomes.

3.4 Curiosity as personality trait

In contrast to momentary states, curiosity has also been conceptualized as a stable personality trait—an enduring individual characteristic that predisposes people toward exploratory behavior and information-seeking across diverse contexts and time periods. Trait curiosity reflects consistent individual differences in the tendency to seek out novel experiences, ask questions, and pursue understanding, functioning as a relatively stable disposition that shapes how people interact with their environment. Within personality psychology, curiosity has been linked to the Big Five personality framework, showing strongest associations with Openness to Experience (McCrae & Costa, 1997; Kashdan et al., 2018). Some researchers argue that curiosity represents a distinct facet within Openness, while others propose it as a separate higher-order trait that cuts across multiple personality dimensions, including aspects of Extraversion (social curiosity) and Conscientiousness (systematic knowledge-seeking).

Contemporary research has revealed curiosity's multidimensional structure through sophisticated measurement approaches. The Five-Dimensional Curiosity Scale (5DC; Kashdan et al., 2018) identifies five distinct factors: Joyous Exploration (enjoyment of novel stimuli), Deprivation Sensitivity (need to resolve knowledge gaps), Stress Tolerance (ability to cope with uncertainty), Social Curiosity (interest in others' behavior), and Thrill Seeking (willingness to take risks for novel experiences). This framework builds on earlier distinctions, such as Litman's (2008) I/D-type epistemic curiosity scales that separate interest-driven from deprivation-driven curiosity. Domain-specific measures have also emerged, such as the Science Curiosity Scale (Kahan et al., 2017) and measures of perceptual versus epistemic curiosity (Collins et al., 2004). These varied assessment tools reveal that trait curiosity is not monolithic but encompasses different ways people characteristically engage with uncertainty and novelty. Understanding these individual differences in trait curiosity helps predict academic achievement, creative problem-solving, psychological well-being, and lifelong learning tendencies, making it a valuable construct for educational and occupational contexts.

3.5 Curiosity as a process: a unifying view

Rather than viewing curiosity as a trait or state, the scientific community has been increasingly recognizing it as a dynamic, multi-component process involving diverse psychological mechanisms working in concert to organize exploration and learning. This process view acknowledges that curiosity emerges from the interaction of multiple systems—attentional, motivational, emotional, cognitive, metacognitive, and behavioral—that collectively guide how organisms navigate uncertainty, seek information or invent and select their own goals.

Several theoretical frameworks have advanced this process perspective. The reward-learning framework (Murayama et al., 2019) conceptualizes curiosity as a dynamic cycle where initial awareness of a knowledge gap triggers interest and exploration, leading to knowledge acquisition that generates an intrinsically rewarding experience, which in turn reinforces further curious behavior (e.g. through generating new questions). The PACE framework (Gruber & Ranganath, 2019) corroborates this view by identifying four key components: Prediction

(generating expectations, measuring errors in prediction), Appraisal (evaluating information value), Curiosity state (the subjective experience and neural activation), and Exploration (searching information to resolve prediction error). These components interact in a cycle where outcomes from exploration feed back to influence future predictions and appraisals.

Computational theories provide perhaps the most detailed process accounts, explicitly modeling the mechanisms underlying curiosity-driven behavior (Baldassarre & Mirolli, 2012; Oudeyer, 2018). These models typically distinguish between a primary learning system that builds knowledge or skills, and a metacognitive system that monitors this primary learning process and generates intrinsic rewards to guide exploration. The metacognitive component tracks patterns in the primary system—such as prediction errors, uncertainty levels, or learning rates—and computes intrinsic rewards that motivate exploration toward situations with high learning potential. This dual-system architecture has enabled to articulate mechanisms enabling curiosity to adaptively focus attention on learnable challenges while avoiding both trivial and impossibly complex situations (Oudeyer, 2018).

This process view unifies previously disparate findings by showing how different curiosity manifestations emerge from the same underlying architecture operating in different contexts. Information-gap theories emphasizing uncertainty reduction, optimal arousal theories focusing on novelty-seeking, and learning progress theories highlighting competence development all describe different aspects or operating modes of the same multi-component system. Context, goals, prior knowledge, and individual differences in component strengths determine the manifestation of curiosity in any given situation. Understanding curiosity as a process also explains its multifaceted nature—why someone might show intense epistemic curiosity in academic domains while displaying little social curiosity, or why the same person might shift between exploration strategies depending on environmental demands. This perspective moves beyond simple individual differences to examine how curiosity emerges from the dynamic interplay of cognitive, affective, motivational processes and the contexts in which they happen.

3.6 Computational models of diverse curiosity processes

Computational modeling has emerged as a powerful approach for formalizing and testing theories of curiosity, enabling researchers to precisely specify mechanisms underlying curiosity-driven exploration and learning in humans and other animals. These models serve multiple purposes: they formalize verbal theories into mathematically precise frameworks, simulate how proposed mechanisms account for already observed phenomena-including proposing explanations of the role of curiosity in explaining other aspects of cognition, and generate testable predictions for experimental validation. For instance, computational models enabled to formulate theories on the role of curiosity-driven exploration in self-organizing long term developmental trajectories, e.g. some models successfully reproduced patterns of vocal development (Moulin-Frier et al., 2014), tool use acquisition (Forestier & Oudeyer, 2016), and visual category learning (Twomey & Westermann, 2018) in human infants. Other models were also used to account for exploratory behaviors in other animals like rodents (Gordon et al., 2014). Examples of predictions from these models that have been experimentally confirmed

include predictions from the Learning Progress hypothesis tested in humans (Ten et al., 2021; Poli et al., 2020, 2022; Leonard et al., 2023).

These computational approaches reveal that curiosity encompasses far more than non-instrumental information seeking, identifying a rich diversity of curiosity processes that vary in their goals, mechanisms, and behavioral manifestations. The models typically distinguish between two major categories (Oudeyer and Kaplan, 2007; Mirolli & Baldassarre, 2013): Knowledge-Based Intrinsically Motivated exploration and learning (KB-IM), focused on improving predictions about the world, and Competence-Based Intrinsically Motivated exploration and learning (CB-IM), centered on improving abilities to achieve goals.

Most computational models of curiosity share a common dual-loop architecture. The first is a low-level learning loop where agents interact with their environment, use internal models to predict or control outcomes, measure prediction or control errors, and update their models accordingly. This loop can employ various learning mechanisms, from self-supervised learning for building predictive models to reinforcement learning for acquiring goal-reaching policies. The second is a higher-level metacognitive loop that guides exploration by monitoring the lower-level system and selecting what to explore next. This metacognitive system includes three key components: (i) a metamodel that tracks patterns in the primary learning system, such as where prediction errors or learning progress occur; (ii) computation of intrinsic rewards quantifying the potential interest of different experiments; and (iii) decision mechanisms for selecting next experiments that maximize expected intrinsic rewards, often using reinforcement learning methods (Barto et al., 2004).

The distinction between KB-IM and CB-IM models reflects fundamentally different forms of curiosity. In KB-IM systems, agents select actions to observe their effects and improve predictive models—curiosity drives the acquisition of knowledge about "what leads to what." In CB-IM systems, agents generate their own goals (and more generally their own problems, or their own games), allocate effort toward achieving them, and learn from goal-achievement errors—curiosity drives the acquisition of competencies. This latter form has been termed "autotelic" exploration (Colas et al., 2022), from Greek auto (self) and telos (goal), emphasizing the self-directed nature of goal generation and pursuit (Forestier et al., 2022). This connects to Csikszentmihalyi's (1991) flow theory, where intrinsic motivation emerges from pursuing self-selected challenges matched to one's abilities.

Models vary along two primary dimensions. First, whether they implement knowledge-based or competence-based/autotelic mechanisms fundamentally shapes what is learned and how. Second, the specific intrinsic reward signals used to quantify interestingness of potential experiments to conduct (or probe actions to make) create different exploration patterns. These rewards include novelty (Sutton, 1990; Dayan & Sejnowski, 1996), prediction errors (Marvin & Shohamy, 2016), surprise (Friston et al., 2017; Schwartenbeck et al., 2019), state visitation density (Ostrovski et al., 2017), competence errors (Barto et al., 2004), epistemic uncertainty (Osband et al., 2023), intermediate complexity (Dubey & Griffiths, 2020), learning progress for predictions (Schmidhuber, 1991; Kaplan & Oudeyer, 2007), learning progress for skills in autotelic agents (Oudeyer & Kaplan, 2007; Baranes and Oudeyer, 2013), predictive information

(Martius et al., 2013), and empowerment (Salge et al., 2014). Each reward type produces distinct exploration strategies, from random exploration with novelty rewards to systematic curriculum learning with progress-based rewards, demonstrating how computational approaches illuminate the diverse mechanisms through which curiosity operates.

3.7 Normative vs heuristic theories of curiosity

Theories of curiosity can be broadly categorized into normative and heuristic (or process) theories, which serve complementary roles in understanding curiosity-driven behavior (Ten et al., 2024). Normative theories adopt a "function-first" approach, beginning with a well-motivated computational objective—typically knowledge maximization (Ten et al., 2024)—and deriving mathematically optimal solutions under specified constraints (Griffiths et al., 2010; Dubey & Griffiths, 2020). These theories prescribe how an ideal rational agent should allocate limited resources to maximize learning, establishing normative benchmarks for curiosity. One example of a key insight from normative analyses is that the inverted-U relationship between knowledge and curiosity, observed in many contexts, emerges as an optimal exploration strategy specifically when learning follows S-shaped curves. As Ten et al. (2024) synthesize, when competence grows according to an S-shaped function, the derivative of knowledge gain peaks at intermediate levels, making it optimal to focus on tasks where one has partial knowledge. However, normative theories also reveal that this inverted-U pattern is not universally optimal: in environments with exponential learning curves, optimal exploration should favor novel stimuli unless additional constraints apply, such as when past exposure correlates with future encounter probability (Dubey & Griffiths, 2020).

In contrast, heuristic or process theories employ a "bottom-up" approach, proposing psychologically plausible mechanisms that approximate optimal solutions while accounting for real-world cognitive, energetic and time constraints. These theories describe how curiosity actually emerges from specific and tractable computational processes involving variables like uncertainty, familiarity, or learning progress, rather than prescribing ideal behavior. While normative theories reveal why certain curiosity patterns (like the curious U) might be adaptive from a computational perspective, heuristic theories explain how resource-limited agents might implement approximations of these optimal strategies. This distinction parallels Marr's levels of analysis: normative theories address the computational level (what problem is being solved and why), while heuristic theories address the algorithmic level (how the problem is actually solved). Together, they provide complementary functional and mechanistic accounts of curiosity, with normative theories establishing theoretical ideals and heuristic theories explaining how these ideals might be realized in practice through implementable cognitive mechanisms.

Some theories straddle this divide: for example, the Learning Progress Hypothesis exemplifies both normative and process aspects. It was initially proposed as a computationally efficient mechanism for enabling agents explore efficiently large environments with several challenges such as many unlearnable tasks, or too many learnable tasks given the whole budget of exploration (Oudeyer et al., 2007; Kaplan and Oudeyer, 2007b). Later on, it was shown that some implementations of learning progress maximization can be optimal for important classes

of learning environments (Lopes and Oudeyer, 2012). This dual nature illustrates how computationally tractable heuristics can sometimes achieve near-optimal performance, bridging the gap between ideal normative solutions and practical cognitive mechanisms.

3.8 Neural bases of curiosity

Research into the neural mechanisms underlying curiosity has employed various methodologies to understand how the brain processes and values information-seeking behaviors. In humans, functional magnetic resonance imaging (fMRI) studies have used trivia paradigms, where participants rate their curiosity about questions before receiving answers, revealing brain activation patterns during anticipatory periods (Kang et al., 2009; Gruber et al., 2014). Electroencephalography (EEG) has also emerged as a portable and affordable tool for continuous monitoring of curiosity states, for example leveraging information in theta oscillations (Begus and Bonawitz, 2020), or combining multiple frequency bands using machine learning (Appriou et al., 2020). Functional near-infrared spectroscopy (fNIRS) has also been employed, detecting hemodynamic responses in orbitofrontal cortex regions using blurred picture paradigms to induce perceptual uncertainty—a state closely related to curiosity (Korniluk et al., 2025). In non-human primates, researchers have utilized observing paradigms, where monkeys choose between informative and uninformative cues that predict rewards, combined with single-unit recordings to examine neuronal responses (Bromberg-Martin & Hikosaka, 2009; Blanchard et al., 2015).

These multimodal investigations have revealed that curiosity activates brain regions traditionally associated with reward processing. In humans, high curiosity states elicit increased activity in the caudate nucleus, bilateral inferior frontal gyrus, putamen, and globus pallidus during the anticipation of information, detectable through fMRI (Kang et al., 2009). EEG studies have shown enhanced theta oscillations in frontal regions during curiosity states, suggesting their potential as real-time markers for monitoring information-seeking behaviors (Begus & Bonawitz, 2020). Similarly, the substantia nigra/ventral tegmental area (SN/VTA) complex and nucleus accumbens show enhanced activation when individuals expect to receive answers to questions they find intriguing (Gruber et al., 2014). Neuronal recordings in monkeys have demonstrated that midbrain dopaminergic cells encode the anticipation of obtaining reliable information, exhibiting stronger excitatory responses when expecting informative versus uninformative cues (Bromberg-Martin & Hikosaka, 2009).

Overall, a crucial finding across species is that behaviours motivated by curiosity and by extrinsic rewards (like food or money) are processed in a convergent manner through a two-level neural organization. At the subcortical level, the dopaminergic reward system treats both types of rewards similarly: midbrain dopamine neurons signal value for both information and material gains, providing a "common currency" that allows direct comparison between them. This supports the mechanisms leading humans to sometimes sacrifice money or time to satisfy curiosity (Kang et al., 2009), and leading monkeys to choose informative cues even when this may be at the cost of juice rewards (Blanchard et al., 2015). However, at the cortical level, the brain maintains a distinction: the orbitofrontal cortex uses separate neural populations to

compute information value versus extrinsic reward value (Blanchard et al., 2015). This makes functional sense—while we need to compare knowledge and food on a single scale for decision-making (hence the shared dopaminergic signal), calculating their respective values requires fundamentally different computations. Information value depends on semantic and epistemic factors (what does this knowledge mean?), whereas extrinsic reward value depends on physiological needs or material benefits (how nutritious is this food? how valuable is this money?). This two-level architecture—unified comparison at the dopaminergic level, specialized computation at the cortical level—provides the neurobiological basis for how organisms can flexibly choose between satisfying curiosity or pursuing material rewards.

3.9 Proximal functionalities of curiosity

Proximal functionalities refer to the immediate, direct benefits that curiosity provides to individual organisms during their lifetime, as opposed to distal evolutionary advantages that emerge over longer timescales across populations (Singh et al., 2010). At the level of individuals, computational models have revealed that curiosity-driven processes enable remarkably efficient learning under resource constraints and challenging environmental conditions (Baldassarre and Mirolli, 2013; Oudeyer, 2018). These models demonstrate that curiosity mechanisms automatically self-organize effective learning curricula, directing organisms toward activities that provide optimal learning opportunities while avoiding tasks that are either too trivial or impossibly difficult (Kaplan & Oudeyer, 2007b), focusing on those that are learnable (Gerken et al., 2011). Rather than requiring external guidance, curiosity-driven systems naturally sequence their exploration from simple to increasingly complex challenges, maximizing learning efficiency within the limited time and energy budgets that characterize real-world learning (Twomey and Westermann, 2018, Dubey and Griffiths, 2020; Ten et al., 2024).

A fundamental proximal function of curiosity is enabling organisms to efficiently acquire comprehensive world models—internal representations of how the environment works and how actions produce outcomes. Through systematic exploration driven by intrinsic motivation, organisms build predictive models that capture environmental regularities, causal relationships, and the consequences of different behaviors (Sutton, 1990; Schmidhuber, 1991; Dayan and Sejnowski, 1996; Gopnik et al., 2004; Friston et al., 2017). These world models prove invaluable when organisms later encounter situations requiring the discovery of rare rewards in uncertain and changing environments.

Beyond acquiring knowledge about the world, curiosity also drives the continuous expansion of diverse skill repertoires, serving as an engine of open-ended learning throughout an individual's lifetime (Singh et al., 2004; Forestier et al. 2022; Colas et al., 2022). Rather than being limited to solving predetermined problems, curiosity-driven exploration enables organisms to make unexpected discoveries and develop competencies whose utility may only become apparent much later. This process creates a growing toolkit of behavioral skills and problem-solving strategies that can be flexibly recombined and reused when facing novel challenges—whether these challenges are externally imposed by environmental demands or internally generated through further curiosity-driven exploration. This open-ended dimension of curiosity enables

individuals to continuously expand their adaptive potential, building capabilities that extend far beyond immediate survival needs and creating the foundation for lifelong learning and innovation (Oudeyer & Smith, 2016; Chu and Schultz, 2020).

Particularly crucial is curiosity's role in solving problems where successful outcomes are extremely rare—a ubiquitous challenge in natural environments (Lehman & Stanley, 2011; Bellemarre et al., 2016). Many real-world problems, such as finding food in unfamiliar territories or developing complex skills like tool use, involve situations where most attempts fail and positive feedback occurs infrequently (Forestier et al., 2022). Conventional trial-and-error learning gets stuck in unproductive patterns when success is so uncommon that learners receive little guidance about which approaches might work. Direct, goal-focused learning strategies that concentrate solely on achieving specific objectives often fail because learners receive insufficient feedback about which directions might prove fruitful. In contrast, curiosity-driven exploration enables organisms to efficiently learn about environmental structure and contingencies through intrinsically motivated investigation, substantially increasing the likelihood of eventually discovering solutions to extrinsic problems (Oudeyer, 2018; Pathak et al., 2017). By fostering exploration of diverse behavioral skills and environmental regularities independent of immediate external rewards, curiosity creates learning pathways toward competencies that would remain unreachable through direct goal pursuit alone (Singh et al., 2004).

3.10 Distal functionalities of curiosity: evolutionary perspective

Beyond immediate benefits to individual learning, curiosity serves distal functions that emerge across evolutionary timescales, both explaining how curiosity itself evolved and demonstrating how it continues to drive evolutionary processes. Singh et al. (2010) provide a computational perspective illustrating how intrinsic motivation systems can evolve because they maximize long-term evolutionary fitness under rapidly changing environmental conditions. Their computer simulations demonstrate that when environments change faster than genetic adaptation can track, and when they contain sparse or deceptive rewards—where focusing solely on external goals leads organisms to get trapped in local optima or fail to discover solutions entirely—intrinsic motivation systems that reward exploration and learning progress achieve superior long-term evolutionary fitness than domain-specific behavioral adaptations. Crucially, curiosity-driven exploration enables organisms to build comprehensive world models and diverse skill repertoires during periods when extrinsic rewards are unavailable, creating a repository of knowledge and capabilities that proves invaluable when survival challenges do arise. This provides a theoretical evolutionary account for why curiosity-driven mechanisms would be favored by natural selection, particularly in species like humans that inhabit rapidly changing social and cultural environments.

Curiosity also functions as an engine of evolutionary innovation through what Oudeyer and Smith (2016) term "curiosity-driven developmental processes." These mechanisms create emergent developmental structures that serve as a reservoir of behavioral and cognitive innovations, which can later be recruited for functions not initially anticipated—a process known as exaptation (Gould, 1991). For instance, computational models suggest that generic

curiosity-driven exploration of vocal capabilities could have spontaneously bootstrapped vocal structures at both individual and population levels, which were subsequently co-opted for language functions through evolutionary processes (Oudeyer, 2006; Moulin-Frier et al., 2014). This illustrates how curiosity can generate the raw material for evolutionary innovation by producing diverse behavioral repertoires whose ultimate utility only becomes apparent over longer timescales.

At the cultural evolution level, Chu et al. (2024) argue that humans' capacity for pursuing seemingly arbitrary or "foolish" goals creates a powerful mechanism for generating and transmitting innovations across generations. Because goals and their associated solutions can be decoupled from original motivations and transmitted culturally, even activities initially pursued purely for intrinsic satisfaction can ultimately yield transformative technologies and knowledge systems. This cultural evolutionary process, mediated by curiosity's drive toward flexible goal-setting and exploration, may represent one of the core features of human cognitive evolution, enabling our species to rapidly and open-endedly accumulate innovations that would be impossible for any individual to discover within a single lifetime.

4. Open Questions and Recent Scientific Developments

4.1 Domains and timescales of curiosity

Curiosity manifests across multiple timescales and domains, from millisecond-level orienting responses to lifelong intellectual pursuits. Understanding this diversity requires addressing an important terminological issue in the literature: some researchers, e.g. Hidi and Renninger (2019), distinguish between "curiosity" (conceived narrowly as short-term information seeking to close knowledge gaps) and "interest" (conceived as a longer-term motivational disposition that develops through phases). However, in this review—consistent with broader usage in cognitive science, neuroscience, and artificial intelligence—we adopt a more encompassing conceptualization of curiosity as the set of processes that drive spontaneous exploration and learning across all timescales (see also Murayama et al., 2019). Under this framework, what Hidi and Renninger call "interest development" represents particular manifestations of curiosity processes operating over extended temporal scales, rather than a fundamentally distinct phenomenon.

Short timescales (seconds to minutes) encompass perceptual curiosity—rapid attention shifts toward novel, surprising, or ambiguous sensory stimuli. This form appears across many species and emerges early in development as infants visually track unexpected events or explore objects through looking, touching, and mouthing. These brief exploratory episodes typically resolve quickly through direct sensory experience but can cascade into longer-duration investigation.

Medium timescales (minutes to hours) involve epistemic curiosity—systematic information-seeking to resolve knowledge gaps or understand causal relationships. This

includes question-asking, problem-solving activities, and exploration of abstract concepts. Even non-human primates demonstrate this form, willingly sacrificing rewards to gain information about uncertain outcomes with no practical value. Social curiosity also operates at these timescales, as individuals investigate others' thoughts, feelings, and behaviors through observation and interaction.

Long timescales (months to years) encompass sustained intellectual curiosity and domain-specific interests—from children's deep fascination with dinosaurs or trains to scientists' lifelong research programs. These extended curiosity-driven pursuits may span decades, involving systematic skill development and knowledge accumulation within specialized domains like mathematics, music, or nature.

The relationship between timescales and domains remains poorly understood. Computational approaches suggest that while core curiosity mechanisms (such as learning progress maximization) may be domain-general and operate across all timescales, their expression becomes increasingly domain-specific as knowledge accumulates and expertise develops within particular areas. This temporal specialization may explain why individuals show varying curiosity patterns across different domains and life periods.

4.2 Which intrinsic rewards explain best human curiosity-driven exploration?

Computational theories have proposed numerous intrinsic rewards that might drive curiosity-based exploration, including novelty (Sutton, 1990; Dayan & Sejnowski, 1996; Gordon et al., 2014), prediction errors (Marvin & Shohamy, 2016), surprise (Friston et al., 2017; Schwartenbeck et al., 2019), empowerment (Salge et al., 2014), epistemic uncertainty (Osband et al., 2023), intermediate complexity (Dubey & Griffiths, 2020), learning progress for predictions (Schmidhuber, 1991; Kaplan & Oudeyer, 2007b), learning progress for competence (Oudeyer & Kaplan, 2007; Baranes & Oudeyer, 2013), and predictive information (Martius et al., 2013). Recent experimental paradigms confirm that humans employ several of these mechanisms, often in combination, to organize their spontaneous exploration (Kobayashi et al., 2019; Ten et al., 2021; Molinaro et al., 2023). This raises fundamental questions: Is there a universal set of intrinsic rewards used by all individuals, or do different people rely on different motivational systems? How are multiple intrinsic rewards integrated? Does this integration depend on context? From a normative perspective, are there universally optimal intrinsic reward systems for exploration? As Ten et al. (2024) demonstrate, the answer to this last question is no—optimal intrinsic rewards depend critically on the statistical properties of the environment being explored, and how they link with the (limited) resources of agents in terms of learning capabilities, energy and time.

Traditional accounts based on novelty, prediction error, or surprise face significant limitations in complex, real-world environments. In large exploration spaces, these mechanisms can trap learners in unlearnable situations—tasks that generate perpetual novelty or surprise but offer no opportunity for meaningful learning. For instance, a child trying to predict truly random events,

or events which cannot be predicted because of lack of available information (e.g. predicting color of the next car appearing in the street), would experience constant surprise without ever improving their predictive model. Similarly, approaches that prioritize exploration of situations with maximum epistemic uncertainty suffer from inefficiency: they direct learners toward the most difficult learnable tasks first, consuming excessive time and cognitive resources before any learning gains materialize, which is problematic when the available time resources are small compared to the space of all learnable tasks in the environment.

The learning progress hypothesis addresses these limitations by proposing that curiosity is driven by the rate of improvement in prediction or control abilities (Kaplan & Oudeyer, 2007; Oudeyer et al., 2007). This mechanism creates a self-organizing curriculum where learners naturally focus on activities that are neither too easy (offering no learning progress) nor too difficult (being currently unlearnable), but rather at an intermediate level of difficulty where learning is maximized. This creates bi-directional interactions between curiosity and learning: curiosity guides attention toward fastly learnable tasks, while the learning that results updates estimates of where future progress is likely, dynamically adjusting the focus of exploration.

Empirical support for learning progress as a key driver of curiosity comes from multiple sources. Computational models implementing learning progress mechanisms successfully reproduce structural regularities in human developmental trajectories, including vocal development (Moulin-Frier et al., 2014), tool use discovery (Forestier & Oudeyer, 2016), and social skill abilities (Doyle et al., 2023). Recent behavioral experiments provide direct evidence that humans monitor learning progress during exploration. Ten et al. (2021) showed that adults who are free to select and explore several learning tasks, show an exploration pattern which is best accounted for by utility functions which combine learning progress and prediction error. Similar findings emerge from studies by Poli et al. (2020, 2022, 2024), Leonard et al. (2023), and Sayalı et al. (2023), collectively demonstrating that learning progress signals guide exploration across diverse tasks and age groups. Critically, children's capacity to use learning progress signals for guiding exploration appears to develop during early and middle childhood: while children ages 6 and older strategically practice more difficult tasks when uncertain about upcoming assessments, younger children show this ability only inconsistently (Serko et al., 2025). Moreover, children's metacognitive predictions about their own learning curves reveal important developmental changes, with 7-8 year-olds predicting gradual improvement on novel skill tasks, while younger children require additional scaffolding to form accurate predictions about how practice leads to mastery (Zhang et al., 2025).

Importantly, the learning progress framework unifies various previous theoretical accounts (Ten et al., 2024). It formalizes long-standing intuitions about "intermediate complexity" or "optimal challenge" without requiring researchers to predefine what constitutes "too easy" or "too difficult"—these categories emerge naturally from the learning dynamics themselves. A task becomes "too easy" when learning progress approaches zero because performance is near ceiling; it becomes "too difficult" when learning progress is zero despite continued effort.

While learning progress appears to be a strong component of human curiosity, the complete picture is more nuanced. Human curiosity systems likely integrate multiple motivational

components, including various intrinsic rewards. The relative weights of these components can shift depending on the situation and context within individuals, and also vary systematically across individuals, contributing to personality differences in curiosity (Molinaro et al., 2023; Kobayashi et al., 2019). For instance, Molinaro et al. (2023) found that children integrate at least three factors in their information-seeking: uncertainty reduction, instrumental utility for action, and hedonic value (preference for positive information). Ten et al. (2021) similarly observed individual differences in how strongly participants weighted learning progress versus other factors in their exploration decisions. These findings suggest that optimal curiosity involves a flexible system that can adapt its motivational weights to match both environmental demands and individual goals, rather than relying on any single intrinsic reward signal.

4.3 Is curiosity always U-shaped? What explains the curious U?

The "curious U" refers to a robust findings in curiosity research: an inverted U-shaped relationship between knowledge and curiosity, where people show peak curiosity for topics about which they have intermediate knowledge or confidence (Ten et al., 2024). This pattern has been demonstrated across numerous experimental paradigms, most notably in trivia question studies where participants rate their curiosity about questions for which they feel moderately confident about knowing the answer (Kang et al., 2009; Baranes et al., 2015). Similar U-shaped patterns emerge when people freely choose among learning tasks or stimuli of varying difficulty, consistently gravitating toward challenges that are neither too easy nor too hard (Kidd et al., 2012; 2014).

However, theoretical analyses reveal that this U-shaped pattern, while common, is not a universal feature of optimal curiosity. Dubey and Griffiths (2020) demonstrated through rational analysis that the shape of the curiosity-knowledge relationship depends critically on environmental structure. When past experience correlates with future encounters—such as when frequently encountered topics are more likely to be relevant again—an inverted U-shape emerges as optimal, supporting complexity-based theories of curiosity. Conversely, when past and future experiences are independent, optimal curiosity should decrease monotonically with knowledge, favoring novelty-seeking instead. Ten et al. (2024) further synthesize multiple theoretical frameworks showing that the U emerges as an optimal solution only under specific assumptions about learning curves and environmental constraints, particularly when competence follows S-shaped growth functions.

These insights suggest that the curious U, while empirically robust in many laboratory settings, reflects the particular environmental structures people typically encounter rather than a fundamental law of curiosity. The prevalence of the U-shape in human behavior may indicate that most real-world learning environments exhibit the statistical properties—such as correlation between past exposure and future relevance—that make intermediate-knowledge targets optimal for exploration. Understanding when and why curiosity deviates from this pattern could provide valuable insights into how environmental structure shapes human information-seeking behavior.

4.4 Links between curiosity, metacognition, agency

The intricate relationships between curiosity, metacognition, and agency represent a critical frontier in understanding how humans regulate their exploratory behavior and learning. Metacognitive skills serve as essential components across the diversity of curiosity processes, enabling individuals to monitor their knowledge states, evaluate learning opportunities, and guide information-seeking behavior effectively (Abdelghani et al., 2023). Recent theoretical work proposes that fundamental aspects of curiosity itself can be understood as specific type(s) of metacognitive feeling that emerges when individuals detect informational needs and assess the potential for acquiring new knowledge (Goupil & Proust, 2023). This metacognitive perspective suggests that curiosity depends on two fundamental evaluative processes: assessing one's current informational state and predicting the likelihood that exploration will yield meaningful learning gains.

The Learning Progress Hypothesis (LPH) exemplifies how metacognitive mechanisms underpin curiosity-driven exploration (Kaplan and Oudeyer, 2007; Oudeyer et al., 2016). This theory assumes that individuals possess metacognitive capacities enabling them to monitor their learning progress, becoming most curious about activities that provide optimal rates of skill or knowledge acquisition. This notion connects closely with the Region of Proximal Learning (RPL) framework (Metcalfe & Kornell, 2005; Son & Metcalfe, 2000), which proposes that people preferentially allocate attention and study time to materials within an optimal learning zone—not too easy to be trivial, nor too difficult to be discouraging. When deciding what to study, individuals rely on metacognitive judgments about their learning state to identify items they believe are "almost known," directing cognitive resources toward this region where learning progress is maximized.". Crucially, in the LPH and RPL perspectives, individuals rely on *subjective* learning progress—the learner's own assessment of their advancement rather than objective external measures: when metacognitive skills are inaccurate or underdeveloped, individuals may thus misjudge their learning progress, potentially leading to suboptimal curiosity allocation. This helps explain why curiosity effectiveness varies considerably across individuals and developmental stages.

Agency represents another fundamental component implicitly assumed by curiosity theories like the LPH. The experience of making learning progress may be insufficient to sustain intrinsic motivation if the individual lacks agency over their exploratory choices. Self-Determination Theory emphasizes that environments supporting autonomy facilitate intrinsic motivation, with experimental evidence showing that external controls can undermine inherently interesting activities (Deci & Ryan, 1985; Deci, 1971). This relationship between agency and curiosity becomes particularly important during late childhood and adolescence, when needs for autonomy intensify and metacognitive abilities continue developing.

Yet, these links are still poorly understood and many open scientific questions remain. For example: How do the co-development trajectories of curiosity, metacognition, and agency interact across different developmental stages? What cognitive load factors influence the deployment of metacognitive skills during curiosity-driven exploration, and how might excessive cognitive demands interfere with effective self-monitoring? Kim et al. (2024) revealed that

people systematically underestimate their own motivation to seek non-instrumental information, suggesting significant gaps in metacognitive awareness about curiosity itself. In addition, recent research has begun exploring whether specific metacognitive skills can be trained to enhance curiosity, with promising preliminary results showing that interventions targeting uncertainty identification, hypothesis generation, and progress assessment can improve children's question-asking behavior and metacognitive efficiency (Abdelghani et al., 2023), but generalization and scaling of such approaches remain to be studied.

4.5 The need for new experimental paradigms for studying open-ended and autotelic free exploration

Besides many advances, contemporary research on curiosity faces a fundamental methodological limitation: most existing experimental paradigms present participants with small, often discrete sets of predetermined learning options or problems to solve. While these controlled approaches have yielded important insights into specific aspects of curiosity-driven behavior, they fail to capture the most distinctive and perhaps most important features of human curiosity—the capacity for open-ended exploration and autonomous goal generation (Chu and Schultz, 2020a, Colas et al., 2022).

The most fascinating questions about curiosity concern how humans and other animals navigate vast spaces of self-generated goals and problems, rather than simply choosing among experimenter-defined alternatives. How do children spontaneously invent new games, devise novel challenges, or create meaningful problems to pursue? How do they leverage language and symbolic reasoning to generate increasingly complex and creative objectives? These processes of autotelic exploration—from Greek *auto* (self) and *telos* (goal)—represent the generative heart of curiosity, where individuals actively construct their own learning landscapes rather than merely exploring within predefined boundaries (Colas et al., 2022; Forestier et al., 2022).

Chu and Schulz (2020b) illuminate this challenge through their work on children's play behavior. They demonstrate that when children are simply asked to "play" versus accomplish a specific task, they fundamentally alter their exploration patterns, often choosing unnecessarily complex paths and creating arbitrary constraints for themselves. This suggests that authentic curiosity-driven exploration involves a sophisticated capacity for goal invention and manipulation that remains poorly understood. Children routinely abandon straightforward solutions in favor of more elaborate, self-imposed challenges—walking in spirals rather than straight lines, or reaching for difficult-to-access objects while ignoring easily available alternatives. These behaviors reveal an intrinsic drive to generate novel problems and explore possibilities beyond immediate instrumental needs. The absence of adequate experimental paradigms for studying such autotelic exploration creates significant gaps in our understanding.

Furthermore, current methodologies cannot address further questions about how people generalize exploration strategies when encountering new domains of self-generated goals. When someone develops expertise in one area of creative problem-solving, how do they transfer those meta-cognitive skills to different domains that may share some similarities in some dimensions?

How do individuals learn to recognize which types of self-generated challenges are likely to be productive versus those that may lead to frustration or dead ends?

Similarly, we lack paradigms for investigating the compositional and linguistic foundations of goal generation. Language appears to play a crucial role in enabling humans to construct increasingly abstract and complex objectives (Vygotsky, 1934), combining simpler concepts into novel challenges. How do children learn to use language not just to communicate about existing goals, but as a generative tool for creating new ones? The capacity to linguistically represent and manipulate abstract goal structures may be fundamental to the open-ended nature of human learning, yet this remains largely unexplored in experimental settings.

Gottlieb and Oudeyer (2018) highlight additional challenges in studying curiosity through traditional neuroscientific approaches. Neuroimaging and laboratory constraints severely limit participants' freedom to engage in natural exploratory behavior, creating a fundamental tension between experimental control and ecological validity. The brain mechanisms underlying genuine curiosity-driven exploration may only emerge during unconstrained, self-directed investigation—precisely the conditions that are difficult to reproduce in controlled laboratory environments.

Furthermore, the temporal dynamics of autotelic exploration pose significant methodological challenges. Real curiosity-driven learning often unfolds over extended periods, with individuals pursuing goals intermittently, abandoning some objectives while returning to others, and gradually developing increasingly sophisticated problem-solving repertoires. Current experimental paradigms typically operate on much shorter timescales and cannot capture these extended developmental processes.

Addressing these limitations requires developing new methodological approaches that can maintain scientific rigor while preserving the essential autonomy of curiosity-driven exploration. This might involve designing digital environments that allow for genuine open-ended exploration while still enabling systematic data collection, or developing longitudinal observational methods that can track natural curiosity-driven learning over extended periods. In addition, new paradigms must grapple with how to study the cultural and social dimensions of goal generation, as human curiosity often emerges through interaction with cultural tools, artifacts, and social expectations that shape what kinds of problems seem worth pursuing.

The development of such paradigms represents more than a methodological challenge—it is essential for understanding curiosity as a foundational mechanism of human learning and creativity. Only by studying how people generate their own learning objectives can we fully comprehend curiosity's role in driving the open-ended cognitive development at the core of human intelligence (see Cognitive Development).

4.6 Development of curiosity across the lifespan

Curiosity behaviours are observable remarkably early in human development, with infants displaying preferential attention to stimuli of moderate novelty—the "Goldilocks effect"—within their first months of life (Kidd et al., 2012). This early perceptual curiosity supports rapid learning about physical properties through visual exploration and object manipulation. Infants can intentionally elicit information from adults through gestures, and these "curiosity bids" enhance learning compared to passively received information (Jirout et al., 2024).

As language develops, children undergo a dramatic shift from perceptual to epistemic curiosity, with "why" questions emerging around ages 2-3 years. This transition reflects developing metacognitive abilities and growing awareness of knowledge gaps. Computational models used to analyze experimental data reveal that while both children and adults use the tracking of expected learning progress to guide curious exploration (Ten et al., 2021; Leonard et al., 2023; Poli et al., 2025), children also rely on uncertainty and surprise cues in ways that differ from adults. In certain contexts, children (ages 5-9) demonstrate heightened curiosity when uncertainty is higher but also when outcomes are less surprising, suggesting reliance on multiple heuristic cues rather than optimal learning indicators (Liquin et al., 2021).

The neural foundations underlying these changes involve gradual maturation of brain networks supporting curiosity. The hippocampus begins functioning early, but connections with the prefrontal cortex—critical for sophisticated appraisal and metacognitive processes that are key for effective curiosity behaviour—become fully mature only around age 13, marking adolescence as a pivotal period for curiosity development (Gruber & Fandakova, 2021). Adolescents gain enhanced capacity for abstract thinking and complex questioning, yet educational research reveals declining academic curiosity during this period, possibly due to educational practices emphasizing extrinsic over intrinsic motivation.

Adult curiosity patterns challenge simple decline narratives. While trait curiosity (general exploratory tendency) shows negative relationships with age, state curiosity (situational information-seeking) demonstrates positive age relationships (Whatley et al., 2025). This suggests older adults become more selective, experiencing heightened curiosity for personally relevant information while showing lower general curiosity traits. Maintaining curiosity in older age predicts better cognitive outcomes and successful aging (Sakaki et al., 2018).

Major questions remain unanswered about curiosity development. The precise mechanisms underlying the shift from child-like to adult-like curiosity triggers require further investigation—does this reflect genuine changes in optimal information-seeking strategies, differences in metacognitive abilities for estimating expected learning progress, or shifts in learning goals from exploration to exploitation? Recent evidence suggests that metacognitive abilities play a crucial role in these developmental transitions: children's capacity to predict how their performance will improve with practice develops gradually, with younger children (ages 4-6) requiring substantial scaffolding to anticipate learning curves that older children (ages 7-8) predict spontaneously (Zhang et al., 2025). Similarly, the ability to strategically allocate practice

based on task difficulty and future uncertainty emerges around age 6, suggesting that effective curiosity-driven exploration depends on developing metacognitive skills for evaluating one's current competencies and predicting future learning (Serko et al., 2025).

Beyond these first studies, longitudinal research is critically needed to track individual trajectories and identify factors that promote or hinder curiosity maintenance across the lifespan. Also, understanding cultural and contextual influences remains limited, with most research conducted in Western populations, limiting generalizability. The relationship between curiosity and other developing cognitive abilities—including executive functions, metacognition, and theory of mind—requires systematic and crosscultural investigation. From a practical perspective, research must also address why academic curiosity often declines during adolescence and how educational practices can better support intrinsic motivation. Similarly, identifying factors that promote curiosity maintenance in older adults could inform interventions supporting successful aging and lifelong learning.

4.7 Long term developmental effects of curiosity: self-organization of developmental trajectories

While numerous computational models of curiosity have been proposed—including those based on novelty, surprise, uncertainty, and information gain—models implementing the Learning Progress Hypothesis have proven particularly productive in revealing how curiosity can self-organize long-term developmental trajectories. These models have uncovered a remarkable property: curiosity-driven exploration based on learning progress spontaneously generates developmental sequences that unfold over extended timescales, creating learning curricula of progressively increasing complexity without any external programming or predetermined developmental schedule (Oudeyer & Kaplan, 2006; Kaplan & Oudeyer, 2007). These mechanisms lead learners to naturally avoid situations that are either trivial or too complex, instead focusing on activities just beyond their current skill level where learning progress is maximized—a zone of proximal challenge that automatically shifts as competencies develop (Oudeyer et al., 2007). This dynamic process results in overlapping waves of developmental phases, where multiple skills develop in parallel but with staggered onsets and peaks, mirroring the cascading nature of human development (Oudeyer & Smith, 2016). These processes also resonates with Vygotsky's zone of proximal development (Vygotsky, 1934) and with the Region of Proximal Learning model (Metcalfe & Kornell, 2005) in educational psychology, which have shown that learners strategically allocate study time based on their metacognitive assessment of which materials fall within this optimal difficulty zone.

The Playground Experiment exemplifies this self-organizing principle, demonstrating how a single curiosity mechanism can drive the emergence of ordered behavioral and cognitive stages resembling those observed in infant development (Oudeyer et al., 2007). In this robotic simulation, an agent equipped only with curiosity-driven exploration and basic learning capabilities spontaneously progressed through distinct developmental phases: first engaging in unorganized body babbling, then systematically exploring individual motor primitives, subsequently discovering object affordances through non-functional then functional

manipulation, and finally achieving social interaction through vocalization with a responsive partner (Kaplan and Oudeyer, 2006), while internally progressively building representations that enable to distinguish self, other objects and other beings (Kaplan and Oudeyer, 2007a). This progression from simple self-exploration to complex social engagement emerged solely from the drive to maximize learning progress, with no preprogrammed developmental sequence.

Vocal development provides particularly compelling evidence for curiosity-driven self-organization. Moulin-Frier et al. (2014) demonstrated how a robot exploring a realistic vocal tract model through learning progress maximization naturally reproduced key stages of infant vocal development. The system first discovered basic phonation control, progressed to producing unarticulated vocal variations, and eventually mastered articulated proto-syllables—a sequence strikingly similar to the progression from reflexive vocalizations to canonical babbling observed in human infants (Oller, 2000). Critically, the model also captured the adaptive transition from self-directed vocal exploration to socially-influenced vocal learning: as the agent's vocal capabilities matured, imitating adult vocalizations began providing higher learning progress than solo exploration, triggering a natural shift toward social learning without any explicit switching mechanism.

Similar developmental cascades emerge in tool use acquisition, where curiosity-driven agents spontaneously discover increasingly sophisticated manipulation strategies that culminate in discovering that speech sounds can be used as tools to influence the behaviours of others, i.e. the discovery of the linguistic functionality of speech (Forestier & Oudeyer, 2016, 2017), recapitulating the major milestones of infant tool development from 6 to 24 months. Recent work has extended further these findings to social domains, showing how virtual infants equipped with curiosity-like intrinsic rewards naturally develop preferences for animate over inanimate stimuli and spontaneously engage in contingent social play when paired with responsive caregivers (Doyle et al., 2023; Kim et al., 2020). Similarly, predictive learning mechanisms based on minimizing prediction error—another form of curiosity-driven exploration—have been shown to generate diverse social cognitive abilities including self-other discrimination, goal-directed action understanding, imitation, and even altruistic helping behaviors (Nagai, 2019; Baraglia et al., 2016). Overall, these works suggest that fundamental social competencies may emerge from domain-general mechanisms of curiosity.

These computational studies converge on a profound insight: the apparent regularity and universality of developmental trajectories need not reflect innate maturational programs, but can instead emerge from the interaction between curiosity-driven learning, embodied constraints, and environmental structure. By automatically generating curricula adapted to current capabilities, curiosity creates a form of developmental canalization where diverse individuals following different specific paths nonetheless traverse similar broad stages—explaining both the universality and individual variability observed in human development.

4.8 How does social interaction influence curiosity behaviours?

Beyond the role of curiosity in leading the emergence of social behaviours in infants, another key question arises at the interface of curiosity and sociality: How does social interaction influence individual curiosity behaviours? Recent empirical work has demonstrated that curiosity can indeed be "contagious," with social cues powerfully shaping what individuals choose to explore and learn about (Dubey et al., 2021). In a series of experiments, participants showed increased curiosity about scientific questions when those questions were presented with high popularity cues (such as a large number of up-votes) compared to low popularity cues. Crucially, this social influence on curiosity was mediated by two key factors: surprise about why others found the question (not-)interesting, and the inferred usefulness of the knowledge gained from answering the question. This finding aligns with further research showing that perceived usefulness is a strong predictor of curiosity, suggesting that social signals help individuals assess the value and importance of potential learning opportunities (Dubey et al., 2022).

The dynamics of curiosity in group learning contexts reveal even more complex social influences. Sinha and colleagues (2017) proposed a theoretical framework distinguishing between individual functions (such as personal knowledge seeking) and interpersonal functions (such as shared exploration and collaborative reasoning) that contribute to curiosity in social settings. Their empirical work demonstrates the strong role interpersonal functions exert on curiosity, highlighting the strong social dimension of knowledge-seeking behaviour. Building on this framework, Paranjape and colleagues (2018) developed computational models that can predict moment-to-moment changes in curiosity based on sequences of behaviours exhibited during group learning. Their findings reveal that convergence to high curiosity across an entire group is most strongly associated with social behaviours such as questions and answers, arguments, sharing findings, and collaborative scientific reasoning including hypothesis generation and justification. These studies collectively demonstrate that curiosity is not only an individual cognitive state but can also emerge through dynamic social interactions, with social cues serving as powerful external influences that can enhance or diminish the intrinsic drive to explore and learn.

In related human-robot interaction research, studies demonstrate that social robots can elicit similar curiosity contagion effects. Gordon and colleagues (2015) showed that children could "catch" curiosity from robots, while Ceha and colleagues (2019) found that robots expressing curiosity through question-asking were perceived as curious and produced contagion effects including increased participant question-asking and hypothesis generation. These findings suggest that social influences on curiosity extend to human-robot interactions, offering potential for educational technologies that leverage social dynamics to enhance learning.

4.9 Curiosity in non-human animals

Scientists have documented various forms of curiosity-driven behaviors across a remarkable diversity of animal species for over a century. Pioneers like Darwin (1878) suggested the existence of curiosity in animals, and subsequent decades have revealed similar behaviors across

birds, fish, rodents, octopuses, bears, and many other taxa (Glickman & Sroges, 1966; Bacon, 1980; Byrne et al., 2002; Heinrich, 1995). However, despite this rich historical foundation, research on animal curiosity has been surprisingly understudied in recent decades, creating significant gaps in our understanding of these fascinating behaviors (Forss et al., 2024). Exceptions include recent neuroscience research which focused extensively on various monkey species, particularly macaques, whose curiosity-driven behaviors have provided crucial insights that have informed our understanding and study of human curiosity and its associated neural circuitry (Monosov, 2024).

Studying curiosity in non-human animals presents unique methodological challenges that researchers must carefully navigate. Animals cannot verbally report their subjective experiences of curiosity, forcing scientists to rely entirely on behavioral observations and physiological measures. Perhaps more fundamentally, animals inhabit vastly different perceptual worlds (umwelten) than humans, meaning that what appears novel or interesting to a human observer may be entirely familiar to the animal, and vice versa (Birchmeier et al., 2023). These limitations require researchers to develop innovative experimental paradigms that can reliably distinguish curiosity-driven behaviors from other motivational states that may also lead to exploratory behaviour such as fear and hunger.

Despite these challenges, researchers have documented striking examples where animals clearly pay costs to obtain non-instrumental information—behaviors that suggest genuine curiosity-like motivation. Harlow's (1950) pioneering work demonstrated that rhesus monkeys would persistently solve complex mechanical puzzles for hours without any external reward, exhibiting what he termed "intrinsically motivated" behavior. More recently, macaque monkeys have been shown to sacrifice water rewards to learn information about gambling outcomes they could not use to change future results, suggesting they valued knowledge for its own sake (Daddaoua et al., 2016; Bromberg-Martin & Monosov, 2020). American black bears in captivity display intense interest in novel objects, even when no food reward is associated (Bacon, 1980). Ravens demonstrate remarkable neophilia, preferentially investigating novel objects thousands of times more than familiar background items, often abandoning these objects after brief exploration once their novelty wears off (Heinrich, 1995). Even fish species like zebrafish show differential responses to novel objects, suggesting that curiosity-like behaviors may be more widespread than previously recognized (Franks et al., 2023).

These findings raise important questions about the cognitive mechanisms underlying animal curiosity, particularly regarding different levels of metacognitive processing. A central debate concerns whether information-seeking behaviors in animals reflect genuine metacognitive awareness—conscious monitoring of one's own knowledge states—or merely demonstrate metacognitive skills without awareness (Carruthers & Williams, 2019). Animals might exhibit sophisticated uncertainty responses and adaptive information-seeking through executive control processes that monitor environmental cues and internal states without necessarily involving explicit awareness of their own cognitive processes.

Beyond questions of metacognition lies the even more complex issue of autotelic exploration. Can some animals engage in truly self-directed autotelic exploration—imagining and setting

their own goals rather than simply responding to environmental stimuli? Do any species demonstrate the capacity to establish learning objectives for themselves, planning future exploratory behaviors based on anticipated knowledge gains? Whether any non-human animals can engage in the kind of self-directed goal generation that characterizes human autotelic exploration remains an open question with important implications for understanding the evolutionary origins of autonomous curiosity.

The study of curiosity across diverse animal species offers invaluable insights for constructing a comprehensive evolutionary account of this fundamental cognitive phenomenon (Ajuwon et al., 2025) . By examining how curiosity manifests across different taxa with varying ecological pressures, cognitive capabilities, and neural architectures, researchers can begin to identify the core features that define curiosity versus those that represent species-specific adaptations (Forss et al., 2024). Such comparative approaches not only illuminate the deep evolutionary roots of curiosity but also provide crucial constraints for theoretical models, helping distinguish universal principles from human-specific elaborations. Understanding animal curiosity may ultimately reveal whether this remarkable drive to seek knowledge represents a fundamental feature of intelligent life or a more recent evolutionary innovation that reaches its full expression only in humans.

5. Broader connections

5.1 Curiosity and education

Curiosity plays a fundamental role in children's learning and development, particularly within educational contexts. Research demonstrates that curiosity enhances learning through multiple mechanisms: experimental studies show how curiosity states improve memory encoding through increased attention, while also enhancing memory consolidation via dopaminergic neuromodulation of the hippocampus (Kang et al., 2009; Gruber et al., 2014; Marvin & Shohamy, 2016). Large-scale longitudinal studies reveal that early childhood curiosity independently predicts better reading and math achievement at kindergarten, with particularly strong effects for children from lower socioeconomic backgrounds (Shah et al., 2018). More broadly and across development, curiosity correlates with long-term academic and professional success, as intellectual curiosity emerges as a "third pillar" of academic performance alongside intelligence and conscientiousness (Von Stumm et al., 2011), and relates to enhanced well-being and life satisfaction (Kashdan & Steger, 2007).

Historically, educational approaches have varied dramatically in their treatment of curiosity. Progressive educators like Maria Montessori and Friedrich Froebel developed child-centered approaches that prioritized open-ended exploration and discovery learning, where children actively pursue their interests and ask curious questions while teachers scaffold challenges of increasing complexity (Oudeyer et al., 2016). These contrast sharply with more traditional, instruction-focused approaches that emphasize direct and top-down knowledge transmission, in a context where questions are most often formulated by teachers to evaluate children’s

knowledge. Modern research validates many principles underlying curiosity-promoting educational environments, showing that learners benefit when they can explore personally meaningful activities that support their psychological needs for autonomy, competence, and relatedness—the three basic needs identified by Self-Determination Theory (Deci & Ryan, 1985).

The relationship between intrinsic and extrinsic motivation in classrooms remains complex and debated. Self-Determination Theory (Deci & Ryan, 1985) distinguishes intrinsic motivation—engaging in activities for their inherent satisfaction—from extrinsic motivation driven by separable outcomes like grades or rewards. While curiosity represents a prototypical form of intrinsic motivation, research reveals that external rewards can sometimes undermine intrinsic motivation for inherently interesting activities (Deci et al., 2001). However, when external structures provide informational feedback about competence rather than controlling behavior, they may support rather than hinder curiosity-driven learning.

Despite curiosity's documented benefits, many contemporary Western schools struggle to provide adequate space for children's expression and development of inquisitiveness (Engel, 2011; Evans et al., 2023). Multiple factors contribute to this challenge, including standardized curricula, time pressures, and classroom management concerns. Critically, this limitation may exacerbate educational inequalities: children from lower socioeconomic backgrounds often arrive at school with different knowledge foundations than teachers expect, and opportunities for curious questioning allow both children to express their knowledge gaps and teachers to recognize them, enabling more personalized instruction (Goudeau et al., 2024). Research demonstrates systematic disparities in classroom participation, with working-class students receiving fewer opportunities to express themselves and less satisfactory responses to their questions compared to their more privileged peers (Goudeau et al., 2023). Moreover, teachers' beliefs about students' capabilities are influenced by socioeconomic status and ethnicity, affecting academic evaluations and recommendations (Doyle et al., 2023). These patterns are compounded by stereotype threat effects, where students from lower socioeconomic backgrounds may self-censor their questions to avoid confirming negative stereotypes about their academic abilities (Croizet & Claire, 1998), while social-class stereotypes more broadly function to maintain educational inequalities (Durante & Fiske, 2017). In addition, children from working-class families often lack the cultural capital needed to ask questions in ways that are valued by schools, despite being the students who could most benefit from curiosity-driven interactions (Calarco, 2011). Classrooms with limited space for curious questions thus create additional barriers for students who would most benefit from adaptive teaching approaches.

Understanding curiosity's mechanisms offers promising directions for educational improvement. Research emphasizes the importance of creating classroom environments where curious question-asking is explicitly encouraged and where children receive satisfying answers to their inquiries (Post & Walma van der Molen, 2019). Recent advances demonstrate that curiosity can be enhanced through targeted interventions: training programs that develop metacognitive skills (Proust et al., 2025), such as identifying uncertainties, formulating hypotheses, conducting organized information searches, and monitoring learning progress, significantly improve children's curiosity-driven behaviors and learning outcomes (Abdelghani et al., 2023). Also, adaptive educational technologies can leverage computational models like the

Learning Progress Hypothesis to provide personalized learning experiences that maintain optimal challenge levels and foster sustained engagement (Clément et al., 2024).

The scientific question of whether curiosity can be trained continues to generate important research. Studies demonstrate that certain forms of curiosity-related skills can indeed be developed through explicit instruction and practice. For instance, interventions that train curious question asking (Abdelghani et al., 2022) or the full "identify-guess-seek-assess" curiosity cycle show measurable improvements in children's ability to formulate meaningful questions and engage in curiosity-driven exploration (Abdelghani et al., 2023). However, changing deeper attitudes toward curiosity appears to require longer-term interventions with sustained social interaction (Post & Walma van der Molen, 2019). These findings suggest that while behavioral aspects of curiosity can be relatively quickly enhanced, fostering curiosity as a robust and long-term learning disposition and skill requires more comprehensive educational approaches that value and systematically support inquisitive behaviors across the school environment.

5.2 Curiosity in AI and robotics: from solving hard exploration problems to open-ended learning

The quest to build curious artificial agents has emerged from a convergence of multiple research domains, each bringing complementary perspectives on how machines might autonomously explore and learn. Initially developed independently in three distinct fields, artificial curiosity has evolved from simple exploration mechanisms to sophisticated autotelic systems capable of setting their own goals and pursuing open-ended discovery.

The historical origins of curious AI machines trace back to three foundational domains. In developmental robotics, researchers sought to model human curiosity and its crucial role in sensorimotor, cognitive, and social development in robotic systems (Barto et al., 2004; Oudeyer et al., 2007; Baldassarre et al., 2014). Their primary objective was understanding how curiosity drives infant development and implementing these principles in artificial agents. Simultaneously, the machine learning community approached curiosity from two angles: in reinforcement learning, where it helped solve problems with sparse rewards or local optima (Andreae & Andreae, 1978; Sutton, 1990; Schmidhuber, 1991), and in supervised learning through active learning algorithms that strategically sample data from unexplored regions or areas of high model uncertainty (Cohn et al., 1994; Thrun, 1995). Meanwhile, evolutionary computation researchers developed "novelty search" algorithms to discover diverse solutions and overcome deceptive fitness landscapes (Lehman & Stanley, 2011; Pugh et al., 2016), prioritizing behavioral diversity over pure performance optimization.

The mechanisms implemented in these artificial systems initially focused on knowledge-based forms of curiosity, with agents driven to explore areas offering maximal novelty (Sutton, 1990; Dayan & Sejnowski, 1996; Gordon et al., 2014), surprise (Friston et al., 2017; Schwartenbeck et al., 2019), information gain and epistemic uncertainty (Osband et al., 2023), empowerment (Salge et al., 2014), or learning progress in predicting action outcomes (Schmidhuber, 1991; Kaplan & Oudeyer, 2007a). A significant evolution occurred with the emergence of autotelic

curiosity—first in robots (Baranes and Oudeyer, 2013; Srivastava et al., 2013; Santucci et al., 2016; Forestier et al., 2022,), then in deep reinforcement learning systems (Colas et al., 2022). These autotelic agents can learn to represent, generate, sample, and pursue their own goals, potentially prioritizing those that maximize learning progress. Notably, quality-diversity algorithms represent a large subfamily of autotelic curiosity approaches, systematically exploring the space of possible behaviors while optimizing performance within each behavioral niche (Pugh et al., 2016; Forestier et al., 2022).

The purposes for implementing curiosity in machines have proven remarkably diverse. First, curiosity serves as a powerful tool for solving hard-exploration problems in reinforcement learning—environments where rewards are extremely sparse or deceptive. This approach enabled breakthrough performances in notoriously difficult video game benchmarks like Montezuma's Revenge (Bellemare et al., 2016; Ecoffet et al., 2021) and complex robotic manipulation tasks (Forestier et al., 2022). Second, curiosity drives automatic curriculum learning, where AI systems progressively tackle tasks of increasing complexity (Narvekar et al., 2020; Portelas et al., 2021), including approaches like unsupervised environment design (Wang et al., 2019, Dennis et al., 2020). Third, curiosity acts as an engine for open-ended learning, enabling agents to continuously acquire novel knowledge and skills of increasing complexity without predefined objectives (Oudeyer et al., 2007; Santucci et al., 2020; Sigaud et al., 2023; Jiang et al., 2023). Finally, curiosity mechanisms help discover diverse learnable behaviors in complex systems, from automatically generating challenging benchmarks for testing frontier AI models (Pourcel et al., 2024; Lu et al., 2025) to assisting scientists in exploring biological networks (Etcheverry et al., 2024), self-organized physical (Falk et al., 2024) or chemical systems (Grizou et al., 2020), and helping artists discover creative patterns (Secretan et al., 2011; Chan et al., 2016).

Recent advances have integrated foundation models with autotelic curiosity, creating agents that leverage vast prior knowledge encoded in large language models. These systems use language as a powerful abstraction for goal representation, enabling agents to generate abstract, culturally-aligned goals that resonate with human interests (Colas et al., 2022; Zhang et al., 2023). Projects like Voyager demonstrate how LLM-powered agents can explore virtual worlds autonomously, building ever-growing libraries of skills through self-driven discovery (Wang et al., 2023). Other projects have explored how autotelic generative AI can continuously self-improve by creating their own problems, learning to solve them and checking the solutions with the help of external tools (e.g. code interpreters or automatic theorem provers), such as self-improving skills for coding (Pourcel et al., 2024) or mathematics (Zhao et al., 2025). Such approaches can also be seen from the applicative perspective of using LLM-based autotelic curiosity for automated scientific discovery, e.g. for generating novel theorems and associated proofs. In this line, the AI Scientist (Lu et al., 2024) demonstrates how AI systems can generate scientific hypotheses, design experiments, analyze results, and even write scientific papers autonomously.

However, implementing autotelic curious agents raises critical safety and ethical considerations. As AI systems approach and potentially exceed human capabilities in some tasks and in some contexts, unchecked agency poses major risks—from deception and pursuit of self-preservation

goals to potential loss of human control (Bengio et al., 2025). These risks emerge particularly when AI systems optimize for rewards without truly understanding the causal structure of the world or human intentions behind those rewards.

Intriguingly, algorithms for autotelic curiosity-driven exploration, when used as tools for assisting human scientists, may help uncover and map diverse emergent capabilities of generative AI systems (Lu et al., 2025), or identify diverse attacks or dangerous behaviours (Samvelyan et al., 2024), helping prevent risks and reinforce safety.

In another perspective, implementing in machines metacognitive skills inherent to curiosity may offer a potential path toward safer AI. As Johnson (2022) and Johnson et al. (2024) argue, metacognition—the set of abilities to monitor, understand and control one's own cognitive processes— may serve as a crucial safeguard mechanism, enabling AI systems to achieve self-awareness about their limitations and potential failure modes. Metacognitive capabilities may allow for self-diagnosis, anomaly detection, and the recognition of situations where the system's knowledge or capabilities are insufficient-what one may call intellectual humility, and a key part of curiosity processes.

Another way in which curiosity-driven AI may potentially contribute to building safer systems is through the key role of curiosity to enable machines to understand causal relationships in their environment, and thus develop deeper comprehension of the world and human values than systems merely trained to maximize predefined objectives. An agent that has explored and understood causal structures through self-directed experimentation would be less likely to engage in reward hacking—such as a cleaning robot hiding dust under a carpet rather than truly cleaning—because it would grasp the underlying human intention behind the task. Indeed, more generally it has been argued that AI systems that learn causal models and maintain calibrated uncertainty about predictions could prevent the brittle and dangerous behaviors (Bengio et al., 2025). In this line, curiosity-driven systems that learn causal models through interaction with humans offer a potential path toward more robust alignment. Rather than imposing rigid moral constraints, these systems could be progressively educated within a particular human culture—similar to how children learn social norms of their social environment—developing an understanding of human values grounded in causal comprehension of social dynamics (Sigaud et al., 2022; Colas et al., 2022).

The study of artificial curiosity ultimately illuminates the evolutionary advantages of curiosity in biological systems. The remarkable effectiveness of curiosity-driven exploration in enabling machines to thrive in complex, changing environments with sparse feedback strongly suggests why similar mechanisms evolved across diverse animal species throughout phylogenetic history. As we build increasingly sophisticated curious machines, we gain deeper insights into one of nature's most powerful learning principles—the drive to explore, discover, and understand our world not for external rewards, but for the intrinsic satisfaction of learning itself.

6. Acknowledgements

This article benefitted from numerous discussions with researchers studying curiosity and its broader connections, including (by alphabetical order) Rania Abdelghani, Gianluca Baldassarre, Andrew Barto, Marc Bellemare, Elizabeth Bonawitz, Cédric Colas, Peter Dayan, Goren Gordon, Jacqueline Gottlieb, Louise Goupil, Mathias Gruber, Frédéric Kaplan, Céleste Kidd, Edith Law, Manuel Lopes, Clément Moulin-Frier, Kou Murayama, Daniel Polani, Francesco Poli, Hélène Sauzéon, Laura Schulz, Olivier Sigaud, Alexander Ten, and members of the Flowers AI & CogSci lab at Inria. I used the Claude AI software to assist in formulating in clear English the explanations and arguments outlined in the chapter. All arguments, explanations, choices of emphasis and citations, and any errors are entirely my own responsibility.

7. Further Readings

Kidd, C., Hayden, B.Y., 2015. The psychology and neuroscience of curiosity. Neuron 88 (3), 449–460.

Murayama, K., FitzGibbon, L., & Sakaki, M. (2019). Process account of curiosity and interest: A reward-learning perspective. Educational Psychology Review, 31, 875-895.

Gottlieb, J., & Oudeyer, P. Y. (2018). Towards a neuroscience of active sampling and curiosity. Nature Reviews Neuroscience, 19(12), 758-770.

Oudeyer, P. Y., Gottlieb, J., & Lopes, M. (2016). Intrinsic motivation, curiosity, and learning: Theory and applications in educational technologies. Progress in brain research, 229, 257-284.

8. References

Abdelghani, R., Oudeyer, P. Y., Law, E., de Vulpillières, C., & Sauzéon, H. (2022). Conversational agents for fostering curiosity-driven learning in children. *International Journal of Human-Computer Studies*, *167*, 102887.

Abdelghani, R., Law, E., Desvaux, C., Oudeyer, P. Y., & Sauzéon, H. (2023, June). Interactive environments for training children's curiosity through the practice of metacognitive skills: a pilot study. In *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference* (pp. 495-501).

Andreae, P. M. & Andreae, J. H. (1978). A teachable machine in the real world. *International Journal of Man-Machine Studies*, 10(3), 301-312.

Appriou, A., Ceha, J., Pramij, S., Dutartre, D., Law, E., Oudeyer, P. Y., & Lotte, F. (2020). Towards measuring states of epistemic curiosity through electroencephalographic signals. In

2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC) (pp. 3228-3235). IEEE.

Ajuwon, V., Monteiro, T., Schnell, A. K., & Clayton, N. S. (2025). To know or not to know? Curiosity and the value of prospective information in animals. *Learning & Behavior*, *53*(1), 114-127.

Bacon, E. S. (1980). Curiosity in the American black bear. *Ursus*, *4*, 153-157.

Baldassarre, G., & Mirolli, M. (2013a). Intrinsically motivated learning systems: An overview. *Intrinsically motivated learning in natural and artificial systems*, 1-14.

Baldassarre, G., & Mirolli, M. (2013b). *Intrinsically motivated learning in natural and artificial systems*. Springer.

Baldassarre, G., Stafford, T., Mirolli, M., Redgrave, P., Ryan, R. M., & Barto, A. (2014). Intrinsic motivations and open-ended development in animals, humans, and robots: An overview. *Frontiers in Psychology*, 5, 985.

Ball, P. (2013). *Curiosity: How science became interested in everything*. University of Chicago Press.

Baraglia, J., Nagai, Y., & Asada, M. (2016). Emergence of altruistic behavior through the minimization of prediction error. *IEEE Transactions on Cognitive and Developmental Systems*, *8*(3), 141-151.

Baranes, A., & Oudeyer, P. Y. (2013). Active learning of inverse models with intrinsically motivated goal exploration in robots. *Robotics and Autonomous Systems*, *61*(1), 49-73.

Baranes, A., Oudeyer, P. Y., & Gottlieb, J. (2015). Eye movements reveal epistemic curiosity in human observers. *Vision Research*, *117*, 81-90.

Barto, A. G., Singh, S., & Chentanez, N. (2004). Intrinsically motivated learning of hierarchical collections of skills. In *Proceedings of the 3rd International Conference on Development and Learning* (pp. 112-119).

Baranes, A., Oudeyer, P. Y., & Gottlieb, J. (2015). Eye movements reveal epistemic curiosity in human observers. *Vision research*, *117*, 81-90.

Begus, K., & Bonawitz, E. (2020). The rhythm of learning: Theta oscillations as an index of active learning in infancy. *Developmental Cognitive Neuroscience*, *45*, 100810.

Bellemare, M., Srinivasan, S., Ostrovski, G., Schaul, T., Saxton, D., & Munos, R. (2016). Unifying count-based exploration and hashing bonus. In *Advances in Neural Information Processing Systems* (pp. 1471-1479).

Bengio, Y., Cohen, M., Fornasiere, D., Ghosn, J., Greiner, P., MacDermott, M., ... & Williams-King, D. (2025). Superintelligent agents pose catastrophic risks: Can scientist AI offer a safer path?. *arXiv preprint arXiv:2502.15657*.

Bennett, D., Bode, S., Brydevall, M., Warren, H., & Murawski, C. (2016). Intrinsic valuation of information in decision making under uncertainty. *PLoS computational biology*, *12*(7), e1005020.

Berlyne, D. (1960). *Conflict, arousal and curiosity*. McGraw-Hill.

Berlyne, D. (1965). *Structure and direction in thinking*. John Wiley and Sons.

Birchmeier, K., Johnson-Ulrich, L., Stein, J., & Forss, S. (2023). The Role of Umwelt in Animal Curiosity: A Within and Between Species Comparison of Novelty Exploration in Mongooses. Animal Behavior and Cognition, 10(4), 329–354. doi:10.26451/abc.10.04.03.2023

Blanchard, T. C., Hayden, B. Y., & Bromberg-Martin, E. S. (2015). Orbitofrontal cortex uses distinct codes for different choice attributes in decisions motivated by curiosity. *Neuron*, *85*(3), 602-614.

Bromberg-Martin, E. S., & Hikosaka, O. (2009). Midbrain dopamine neurons signal preference for advance information about upcoming rewards. *Neuron*, *63*(1), 119-126.

Bromberg-Martin, E. S., & Monosov, I. E. (2020). Neural circuitry of information seeking. *Current Opinion in Behavioral Sciences*, *35*, 62-70.

Byrne, R. A., Kuba, M., & Griebel, U. (2002). Lateral asymmetry of eye use in Octopus vulgaris. *Animal Behaviour*, *64*(3), 461-468.

Carruthers, Peter and Williams, David M. (2019) Comparative Metacognition. Animal Behavior and Cognition, 6 (4). pp. 278-288. ISSN 2372-5052.

Ceha, J., Chhibber, N., Goh, J., McDonald, C., Oudeyer, P. Y., Kulić, D., & Law, E. (2019, May). Expression of curiosity in social robots: Design, perception, and effects on behaviour. In Proceedings of the 2019 CHI conference on human factors in computing systems (pp. 1-12).

Chan, M. T., Gorbet, R., Beesley, P., & Kulić, D. (2016). Interacting with curious agents: User experience with interactive sculptural systems. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 151-158). IEEE.

Chu, J., & Schulz, L. E. (2020a). Play, curiosity, and cognition. *Annual Review of Developmental Psychology*, *2*(1), 317-343.

Chu, J., & Schulz, L. (2020b). Exploratory play, rational action, and efficient search. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 42).

Chu J, Tenenbaum JB, Schulz LE. In praise of folly: flexible goals and human cognition. Trends Cogn Sci. 2024 Jul;28(7):628-642.

Cohn, D., Ghahramani, Z., & Jordan, M. (1994). Active learning with statistical models. *Advances in Neural Information Processing Systems*, 7.

Colas, C., Karch, T., Sigaud, O., & Oudeyer, P. Y. (2022). Autotelic agents with intrinsically motivated goal-conditioned reinforcement learning: A short survey. *Journal of Artificial Intelligence Research*, *74*, 1159-1199.

Colas, C., Karch, T., Moulin-Frier, C., & Oudeyer, P. Y. (2022). Language and culture internalization for human-like autotelic AI. *Nature Machine Intelligence*, 4(12), 1068-1076.

Collins, R. P., Litman, J. A., & Spielberger, C. D. (2004). The measurement of perceptual curiosity. Personality and Individual Differences, 36(5), 1127-1141.

Csikszentmihalyi, M. (1991). *Flow: The psychology of optimal experience*. Harper Perennial.

Daddaoua, N., Lopes, M., & Gottlieb, J. (2016). Intrinsically motivated oculomotor exploration guided by uncertainty reduction and conditioned reinforcement in non-human primates. *Scientific Reports*, *6*, 20202.

Daston, L., & Park, K. (1998). *Wonders and the order of nature, 1150-1750*. Zone Books.

Dayan, P., & Sejnowski, T. J. (1996). Exploration bonuses and dual control. *Machine Learning*, *25*(1), 5-22.

De Charms, R. (1968). *Personal causation: The internal affective determinants of behavior*. Academic Press.

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, *18*(1), 105-115.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Plenum.

Deci, E. L., Koestner, R., & Ryan, R. M. (2001). Extrinsic rewards and intrinsic motivation in education: Reconsidered once again. *Review of Educational Research*, *71*(1), 1-27.

Dennis, M., Jaques, N., Vinitsky, E., Bayen, A., Russell, S., Critch, A., & Levine, S. (2020). Emergent complexity and zero-shot transfer via unsupervised environment design. *Advances in neural information processing systems*, *33*, 13049-13061.

Doyle, C., Shader, S., Lau, M., Sano, M., Yamins, D., & Haber, N. (2023). Intrinsically motivated social play in virtual infants. In *Intrinsically-Motivated and Open-Ended Learning Workshop @ NeurIPS2023*.

Doyle, C., Shader, S., Lau, M., Sano, M., Yamins, D. L., & Haber, N. (2023). Developmental curiosity and social interaction in virtual agents. Proceedings of CogSci 2023.

Dubey, R., & Griffiths, T. L. (2020). Reconciling novelty and complexity through a rational analysis of curiosity. *Psychological Review*, *127*(3), 455-476.

Dubey, R., Mehta, H., & Lombrozo, T. (2021). Curiosity is contagious: A social influence intervention to induce curiosity. *Cognitive Science*, *45*(2), e12937.

Ecoffet, A., Huizinga, J., Lehman, J., Stanley, K. O., & Clune, J. (2021). First return, then explore. *Nature*, 590(7847), 580-586.

Embrey, J. R., Li, A. X., Liew, S. X., & Newell, B. R. (2024). The effect of noninstrumental information on reward learning. *Memory & Cognition*, *52*(5), 1210-1227.

Engel, S. (2011). Children's need to know: Curiosity in schools. Harvard educational review, 81(4), 625-645.

Etcheverry, M., Moulin-Frier, C., & Oudeyer, P. Y. (2024). Hierarchically organized latent modules for exploratory search in morphogenetic systems. *Advances in Neural Information Processing Systems*, 33, 4846-4859.

Falk, M. J., Roach, F. D., Gilpin, W., & Murugan, A. (2024). Curiosity-driven search for novel nonequilibrium behaviors. *Physical Review Research*, *6*(3), 033052.

Fandakova, Y., & Gruber, M. J. (2021). States of curiosity and interest enhance memory differently in adolescents and in children. *Developmental Science*, *24*(1), e13005.

Fastrich, G. M., Kerr, T., Castel, A. D., & Murayama, K. (2018). The role of interest in memory for trivia questions: An investigation with a large-scale database. Motivation Science, 4(3), 227-250.

Festinger, L. (1957). *A theory of cognitive dissonance*. Row, Peterson.

Forestier, S., & Oudeyer, P. Y. (2016). Overlapping waves in tool use development: a curiosity-driven computational model. In *2016 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)* (pp. 238-245). IEEE.

Forestier, S., & Oudeyer, P. Y. (2017). A unified model of speech and tool use early development. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 39).

Forestier, S., Portelas, R., Mollard, Y., & Oudeyer, P. Y. (2022). Intrinsically motivated goal exploration processes with automatic curriculum learning. *Journal of Machine Learning Research*, *23*(152), 1-41.

Forss, S. I., Willems, E. P., Call, J., & van Schaik, C. P. (2024). A transdisciplinary view on curiosity beyond linguistic humans: Animals, infants, and artificial intelligence. *Biological Reviews*, *99*(1), 1-18.

Franks, B., Gaffney, L. P., Graham, C., & Weary, D. M. (2023). Curiosity in zebrafish (Danio rerio)? Behavioral responses to 30 novel objects. *Frontiers in Veterinary Science*, *9*, 1062420.

Friston, K. J., Lin, M., Frith, C. D., Pezzulo, G., Hobson, J. A., & Ondobaka, S. (2017). Active inference, curiosity and insight. *Neural Computation*, *29*(10), 2633-2683.

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active inference and learning. *Neuroscience & Biobehavioral Reviews*, 68, 862-879.

Gerken, L., Balcomb, F. K., & Minton, J. L. (2011). Infants avoid labouring in vain by attending more to learnable than unlearnable linguistic patterns. *Developmental Science*, 14(5), 972-979.

Glickman, S. E., & Sroges, R. W. (1966). Curiosity in zoo animals. *Behaviour*, *26*(1-2), 151-187.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: causal maps and Bayes nets. *Psychological review*, *111*(1), 3.

Gottlieb, J., & Oudeyer, P. Y. (2018). Towards a neuroscience of active sampling and curiosity. Nature Reviews Neuroscience, 19(12), 758-770.

Gordon, G., Fonio, E., & Ahissar, E. (2014). Emergent exploration via novelty management. *Journal of Neuroscience*, *34*(38), 12646-12661.

Gordon, G., Breazeal, C., Engel, S. (2015) Can children catch curiosity from a social robot? In: Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction. pp. 91–98. ACM.

Grizou, J., Points, L. J., Sharma, A., & Cronin, L. (2020). A curious formulation robot enables the discovery of a novel protocell behavior. Science advances, 6(5), eaay4237.

Gruber, M. J., & Ranganath, C. (2019). How curiosity enhances hippocampus-dependent memory: The prediction, appraisal, curiosity, and exploration (PACE) framework. *Trends in Cognitive Sciences*, *23*(12), 1014-1025.

Gruber, M. J., Gelman, B. D., & Ranganath, C. (2014). States of curiosity modulate hippocampus-dependent learning via the dopaminergic circuit. *Neuron*, *84*(2), 486-496.

Gruber, M. J., & Fandakova, Y. (2021). Curiosity in childhood and adolescence—what can we learn from the brain. *Current Opinion in Behavioral Sciences*, *39*, 178-184.

Gruber, M. J., Valji, A., & Ranganath, C. (2019). Curiosity and learning: A neuroscientific perspective.

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., and Tenenbaum, J. B. (2010). "Probabilistic models of cognition: exploring representations and inductive bi- ases". In: Trends in Cognitive Sciences 14.8, pp. 357–364. doi: 10.1016/j.tics. 2010.05.004.

Gould, S. J. (1991). Exaptation: A crucial tool for an evolutionary psychology. Journal of Social Issues, 47(3), 43-65.

Goupil, L., & Proust, J. (2023). Curiosity as a metacognitive feeling. Cognition, 231, 105325.

Harlow, H. (1950). Learning and satiation of response in intrinsically motivated complex puzzle performances by monkeys. *Journal of Comparative and Physiological Psychology*, *43*, 289-294.

Harrison, P. (2001). Curiosity, forbidden knowledge, and the reformation of natural philosophy in early modern England. *Isis*, *92*(2), 265-290.

Heinrich, B. (1995). Neophilia and exploration in juvenile common ravens, Corvus corax. *Animal Behaviour*, *50*(3), 695-704.

Hidi, S. E., & Renninger, K. A. (2019). Interest development and its relation to curiosity: Needed neuroscientific research. Educational Psychology Review, 31(4), 833-852.

Hull, C. L. (1943). *Principles of behavior: An introduction to behavior theory*. Appleton-Century-Croft.

Hunt, J. M. (1965). Intrinsic motivation and its role in psychological development. *Nebraska Symposium on Motivation*, *13*, 189-282.

James, W. (1891). *The principles of psychology* (Vol. II). Macmillan.

Jiang, M., Rocktäschel, T., & Grefenstette, E. (2023). General intelligence requires rethinking exploration. *Royal Society Open Science*, *10*(6), 230539.

Jirout, J. J., Evans, N. S., & Son, L. K. (2024). Curiosity in children across ages and contexts. *Nature Reviews Psychology*, *3*(9), 622-635.

Jirout, J. J., & Evans, N. (2023). Exploring to learn: Curiosity, breadth and depth of exploration, and recall in young children. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 45, No. 45).

Johnson, S.G., Karimi, A.H., Bengio, Y., Chater, N., Gerstenberg, T., Larson, K., Levine, S., Mitchell, M., Rahwan, I., Schölkopf, B. and Grossmann, I., (2024) Imagining and building wise machines: The centrality of AI metacognition. arXiv preprint arXiv:2411.02478.

Johnson, B. (2022). Metacognition for artificial intelligence system safety–An approach to safe and desired behavior. *Safety Science*, *151*, 105743.

Kagan, J. (1972). Motives and development. *Journal of Personality and Social Psychology*, *22*, 51-66.

Kahan, D. M., Landrum, A., Carpenter, K., Helft, L., & Hall Jamieson, K. (2017). Science curiosity and political information processing. Political Psychology, 38, 179-199.

Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T., & Camerer, C. F. (2009). The wick in the candle of learning: Epistemic curiosity activates reward circuitry and enhances memory. *Psychological Science*, *20*(8), 963-973.

Kaplan, F., & Oudeyer, P. Y. (2007a). The progress drive hypothesis: An interpretation of early imitation. In C. Nehaniv & K. Dautenhahn (Eds.), *Models and Mechanisms of Imitation and Social Learning in Robots, Humans and Animals* (pp. 361-377). Cambridge University Press.

Kaplan, F., & Oudeyer, P. Y. (2007b). In search of the neural circuits of intrinsic motivation. *Frontiers in Neuroscience*, *1*, 17-28.

Kashdan, T. B., Stiksma, M. C., Disabato, D. J., McKnight, P. E., Bekier, J., Kaji, J., & Lazarus, R. (2018). The five-dimensional curiosity scale: Capturing the bandwidth of curiosity and identifying four unique subgroups of curious people. Journal of Research in Personality, 73, 130-149.

Kenny, N. (2004). *The uses of curiosity in early modern France and Germany*. Oxford University Press.

Kidd, C., & Hayden, B. Y. (2015). The psychology and neuroscience of curiosity. *Neuron*, *88*(3), 449-460.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2014). The Goldilocks effect in infant auditory attention. *Child development*, *85*(5), 1795-1804.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS ONE*, *7*(5), e36399.

Kim, S., Sakaki, M., & Murayama, K. (2024). Metacognition of curiosity: People underestimate the seductive lure of non-instrumental information. *Psychonomic Bulletin & Review*, *31*(3), 1233-1244.

Kim, K., Sano, M., De Freitas, J., Haber, N., & Yamins, D. (2020). Active world model learning with progress curiosity. In *International Conference on Machine Learning* (pp. 5306-5315).

Korniluk, A., Gawda, B., Chojak, M., & Gawron, A. (2025). The neural markers of perceptual uncertainty/curiosity—A functional near-infrared spectroscopy pilot study. *Brain Sciences*, *15*(4), 411.

Lehman, J., & Stanley, K. O. (2011). Abandoning objectives: Evolution through the search for novelty alone. *Evolutionary Computation*, 19(2), 189-223.

Leonard, J. A., Cordrey, S. R., Liu, H. Z., & Mackey, A. P. (2023). Young children calibrate effort based on the trajectory of their performance. *Developmental Psychology*, *59*(3), 609-621.

Liquin, E. G., Callaway, F., & Lombrozo, T. (2021). Developmental change in what elicits curiosity. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 43, No. 43).

Litman, J. A. (2008). Interest and deprivation factors of epistemic curiosity. Personality and Individual Differences, 44(7), 1585-1595.

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin*, *116*(1), 75-98.

Lopes, M., & Oudeyer, P. Y. (2012, November). The strategic student approach for life-long exploration and learning. In *2012 IEEE international conference on development and learning and epigenetic robotics (ICDL)* (pp. 1-8). IEEE.

Lu, C., Lu, C., Lange, R. T., Foerster, J., Clune, J., & Ha, D. (2024). The AI scientist: Towards fully automated open-ended scientific discovery. *arXiv preprint arXiv:2408.06292*.

Lu, C., Hu, S., & Clune, J. (2025). Automated capability discovery via foundation model self-exploration. *ICLR 2025*.

Lydon-Staley, D. M., Zhou, D., Blevins, A. S., Zurn, P., & Bassett, D. S. (2021). Hunters, busybodies and the knowledge network building associated with deprivation curiosity. *Nature human behaviour*, *5*(3), 327-336.

Martius, G., Der, R., & Ay, N. (2013). Information driven self-organization of complex robotic behaviors. *PLoS ONE*, *8*(5), e63400.

Marvin, C. B., & Shohamy, D. (2016). Curiosity and reward: Valence predicts choice and information prediction errors enhance learning. *Journal of Experimental Psychology: General*, *145*(3), 266-272.

Metcalfe, J., & Kornell, N. (2005). A region of proximal learning model of study time allocation. Journal of memory and language, 52(4), 463-477.

Mirolli, M., & Baldassarre, G. (2013). Functions and mechanisms of intrinsic motivations: The knowledge versus competence distinction. In *Intrinsically motivated learning in natural and artificial systems* (pp. 49-72).

Monosov, I. E. (2024). Curiosity: primate neural circuits for novelty and information seeking. *Nature Reviews Neuroscience*, *25*(3), 195-208.

Montgomery, K. (1954). The role of exploratory drive in learning. *Journal of Comparative and Physiological Psychology*, *47*, 60-64.

Moulin-Frier, C., Nguyen, S. M., & Oudeyer, P. Y. (2014). Self-organization of early vocal development in infants and machines: the role of intrinsic motivation. *Frontiers in psychology*, *4*, 1006.

Murayama, K., FitzGibbon, L., & Sakaki, M. (2019). Process account of curiosity and interest: A reward-learning perspective. *Educational Psychology Review*, *31*, 875-895.

Nagai, Y. (2019). Predictive learning: its key role in early cognitive development. *Philosophical Transactions of the Royal Society B*, *374*(1771), 20180030.

Narvekar, S., Peng, B., Leonetti, M., Sinapov, J., Taylor, M. E., & Stone, P. (2020). Curriculum learning for reinforcement learning domains: A framework and survey. *Journal of Machine Learning Research*, *21*(181), 1-50.

Oller, D. K. (2000). *The emergence of the speech capacity*. Psychology Press.

Osband, I., Wen, Z., Asghari, S. M., Dwaracherla, V., Ibrahimi, M., Lu, X., & Van Roy, B. (2023). Epistemic neural networks. *Advances in Neural Information Processing Systems*, *36*, 2795-2823.

Ostrovski, G., Bellemare, M. G., Oord, A. V. D., & Munos, R. (2017). Count-based exploration with neural density models. In *Proceedings of the International Conference on Machine Learning*.

Oudeyer, P. Y. (2018). Self-organization in the evolution of speech (2nd edition). Oxford University Press.

Oudeyer, P. Y. (2018). Computational theories of curiosity-driven learning. In G. Gordon (Ed.), *The new science of curiosity*. NOVA.

Oudeyer, P. Y., & Kaplan, F. (2006). Discovering communication. *Connection Science*, *18*(2), 189-206.

Oudeyer, P. Y., & Kaplan, F. (2007). What is intrinsic motivation? A typology of computational approaches. *Frontiers in Neurorobotics*, *1*, 6.

Oudeyer, P. Y., Kaplan, F., & Hafner, V. V. (2007). Intrinsic motivation systems for autonomous mental development. *IEEE Transactions on Evolutionary Computation*, *11*(2), 265-286.

Oudeyer, P.-Y. and Smith, L. (2016). How evolution can work through curiosity-driven developmental process. *Top. Cogn. Sci*, 8(2):492–502.

Oudeyer, P. Y., Gottlieb, J., & Lopes, M. (2016). Intrinsic motivation, curiosity, and learning: Theory and applications in educational technologies. *Progress in brain research*, *229*, 257-284.

Pathak, D., Agrawal, P., Efros, A. A., & Darrell, T. (2017). Curiosity-driven exploration by self-supervised prediction. In *International Conference on Machine Learning* (pp. 2778-2787).

Poli, F., Meyer, M., Mars, R. B., & Hunnius, S. (2022). Contributions of expected learning progress and perceptual novelty to curiosity-driven exploration. *Cognition*, *225*, 105119.

Poli, F., Serino, G., Mars, R. B., & Hunnius, S. (2020). Infants tailor their attention to maximize learning. *Science Advances*, *6*(39), eabb5053.

Poli, F., O'Reilly, J. X., Mars, R. B., & Hunnius, S. (2024). Curiosity and the dynamics of optimal exploration. *Trends in Cognitive Sciences*, *28*(5), 441-453.

Poli, F., Meyer, M., Mars, R. B., & Hunnius, S. (2025). Exploration in 4-year-old children is guided by learning progress and novelty. *Child Development*, *96*(1), 192-202.

Pomata, G., & Siraisi, N. G. (Eds.). (2005). *Historia: Empiricism and erudition in early modern Europe*. MIT Press.

Portelas, R., Colas, C., Weng, L., Hofmann, K., & Oudeyer, P. Y. (2021). Automatic curriculum learning for deep RL: A short survey. In *IJCAI 2020-International Joint Conference on Artificial Intelligence*.

Pourcel, J., Colas, C., Molinaro, G., Oudeyer, P. Y., & Teodorescu, L. (2024). ACES: Generating diverse programming puzzles with autotelic language models and semantic descriptors. *NeurIPS*.

Pugh, J. K., Soros, L. B., & Stanley, K. O. (2016). Quality diversity: A new frontier for evolutionary computation. *Frontiers in Robotics and AI*, 3, 40.

Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of personality and social psychology*, *43*(3), 450.

Salge, C., Glackin, C., & Polani, D. (2014). Changing the environment based on empowerment as intrinsic motivation. *Entropy*, *16*(5), 2789-2819.

Santucci, V. G., Oudeyer, P. Y., Barto, A., & Baldassarre, G. (2020). Intrinsically motivated open-ended learning in autonomous robots. *Frontiers in neurorobotics*, *13*, 115.

Santucci, V. G., Baldassarre, G., & Mirolli, M. (2016). GRAIL: A goal-discovering robotic architecture for intrinsically-motivated learning. *IEEE Transactions on Cognitive and Developmental Systems*, *8*(3), 214-231.

Sakaki, M., Yagi, A., & Murayama, K. (2018). Curiosity in old age: A possible key to achieving adaptive aging. *Neuroscience & Biobehavioral Reviews*, *88*, 106-116.

Samvelyan, M., Raparthy, S.C., Lupu, A., Hambro, E., Markosyan, A., Bhatt, M., Mao, Y., Jiang, M., Parker-Holder, J., Foerster, J. and Rocktäschel, T., (2024) Rainbow teaming: Open-ended generation of diverse adversarial prompts. *Advances in Neural Information Processing Systems*, *37*, pp.69747-69786.

Sayalı, C., Heling, E., & Cools, R. (2023). Learning progress mediates the link between cognitive effort and task engagement. *Cognition*, *236*, 105418.

Schmidhuber, J. (1991). Curious model-building control systems. In Proceedings of the International Joint Conference on Neural Network, volume 2, pages 1458–1463.

Schwartenbeck, P., Passecker, J., Hauser, T. U., FitzGerald, T. H., Kronbichler, M., & Friston, K. J. (2019). Computational mechanisms of curiosity and goal-directed exploration. *eLife*, *8*, e41703.

Secretan, J., Beato, N., D'Ambrosio, D. B., Rodriguez, A., Campbell, A., Folsom-Kovarik, J. T., & Stanley, K. O. (2011). Picbreeder: A case study in collaborative evolutionary exploration of design space. *Evolutionary Computation*, 19(3), 373-403.

Serko, D., Leonard, J., & Ruggeri, A. (2025). Children strategically decide what to practice. Child Development, 96(5), 1619-1631. https://doi.org/10.1111/cdev.14268

Sigaud, O., Akakzia, A., Caselles-Dupré, H., Colas, C., Oudeyer, P. Y., & Chetouani, M. (2022). Toward teachable autotelic agents. *IEEE Transactions on Cognitive and Developmental Systems*, 15(3), 1070-1084.

Sigaud, O., Baldassarre, G., Colas, C., Doncieux, S., Duro, R., Oudeyer, P.Y., Perrin-Gilbert, N. and Santucci, V.G., (2023) A definition of open-ended learning problems for goal-conditioned agents. *arXiv preprint arXiv:2311.00344*.

Singh, S., Lewis, R. L., Barto, A. G., & Sorg, J. (2010). Intrinsically motivated reinforcement learning: An evolutionary perspective. *IEEE Transactions on Autonomous Mental Development*, *2*(2), 70-82.

Singh, S., Barto, A. G., & Chentanez, N. (2004). Intrinsically motivated reinforcement learning. In *Advances in Neural Information Processing Systems* (pp. 1281-1288).

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. Journal of Experimental Psychology: Learning, Memory, and Cognition, 26(1), 204-221.

Srivastava, R. K., Steunebrink, B. R., & Schmidhuber, J. (2013). First experiments with powerplay. *Neural Networks*, *41*, 130-136.

Sutton, R. (1990). Integrated architectures for learning, planning, and reacting based on approximating dynamic programming. In Proceedings of the Seventh International Conference on Machine Learning, pages 216–224.

Ten, A., Kaushik, P., Oudeyer, P. Y., & Gottlieb, J. (2021). Humans monitor learning progress in curiosity-driven exploration. *Nature Communications*, *12*(1), 5972.

Ten, A., Oudeyer, P. Y., Sakaki, M., & Murayama, K. (2024). The curious U: Integrating theories linking knowledge and information-seeking behavior. https://doi.org/10.31234/osf.io/s8mkj

Thrun, S. (1995). Exploration in active learning. *Handbook of Brain Science and Neural Networks*, 381-384.

Twomey, K. E., & Westermann, G. (2018). Curiosity-based learning in infants: A neurocomputational approach. *Developmental Science*, *21*(4), e12629.

Vygotsky, L. S. Thought and Language (MIT press, 1934).

Wang, R., Lehman, J., Clune, J., & Stanley, K. O. (2019). Poet: open-ended coevolution of environments and their optimized solutions. In *Proceedings of the genetic and evolutionary computation conference* (pp. 142-151).

Wang, G., Xie, Y., Jiang, Y., Mandlekar, A., Xiao, C., Zhu, Y., ... & Anandkumar, A. (2023). Voyager: An open-ended embodied agent with large language models. *Transactions on Machine Learning Research*.

Whatley, M. C., Murayama, K., Sakaki, M., & Castel, A. D. (2025). Curiosity across the adult lifespan: Age-related differences in state and trait curiosity. *PloS one*, *20*(5), e0320600.

White, R. (1959). Motivation reconsidered: The concept of competence. *Psychological Review*, *66*, 297-333.

Zhao, A., Wu, Y., Yue, Y., Wu, T., Xu, Q., Yue, Y., Lin, M., Wang, S., Wu, Q., Zheng, Z. and Huang, G., (2025) Absolute Zero: Reinforced Self-play Reasoning with Zero Data. *arXiv e-prints*, pp.arXiv-2505.

Zhang, X., Carrillo, B. A., Christakis, A., & Leonard, J. A. (2025). Children predict improvement on novel skill learning tasks. Child Development, 96(3), 1177-1188. https://doi.org/10.1111/cdev.14232Retry

Zhang, J., Lehman, J., Stanley, K., & Clune, J. (2023) OMNI: Open-endedness via Models of human Notions of Interestingness. In *The Twelfth International Conference on Learning Representations*.