Balkan Journal of Philosophy (2026), vol. 18, issue 1, 32 – 49

UNDERSTANDING BEYOND KNOWLEDGE: FICTION, MODELS, AND COGNITIVE GAIN

Kênio Estrela*

Abstract: This paper examines the epistemic status of understanding and argues that it cannot be adequately reduced to propositional knowledge or to justified true belief. The analysis challenges strictly truth-centered accounts of cognition and defends a conception of understanding grounded in structured interpretive engagement with representational systems. Drawing on recent debates in epistemology and philosophy of science that distinguish understanding from knowledge (e.g., Kvanvig 2003; Grimm 2006; de Regt 2017), the paper situates understanding as a distinct epistemic achievement with its own normative profile, irreducible to belief or information possession. It is argued that scientific models, literary fictions, and contemporary AI-mediated representations systematically rely on idealized or non-literal structures that nonetheless yield genuine epistemic gains. The central claim is that understanding arises through norm-governed practices of interpretation and use, rather than through the mere accumulation of accurate propositions. Scientific models contribute to understanding by isolating explanatory dependencies, fictional narratives afford experiential, perspectival, and meaning-structuring insight, and AI systems support human understanding by facilitating linguistically mediated exploratory and interpretive activities. In each case, departures from literal truth function as epistemically productive constraints rather than as defects. By articulating a unified framework that accounts for these practices, the paper clarifies the relationship between understanding and knowledge and defends the legitimacy of fictional and idealized representations as indispensable components of human sense-making and cognition.

Keywords: understanding, knowledge, scientific models, fiction, representation, epistemology.

1. Introduction

A widespread assumption in epistemology is that genuine cognitive achievement primarily consists in the acquisition of true propositions about the world. On this view, epistemic success is assessed mainly in terms of truth, justification, and reliability, and knowledge is treated as the central epistemic notion. Within such a framework, science appears as the paradigmatic domain of cognition, while art and fiction are commonly regarded as epistemically secondary, associated more with imagination or aesthetic value than with genuine understanding. This assumption underlies many traditional epistemological frameworks, in which understanding is treated either as a derivative notion or as a merely psychological accompaniment to knowledge.

*Escuela de Filosofía, Universidad Finis Terrae, Santiago; kestrela@uft.cl

This orientation reflects what has often been described as a truth-centered or belief-centered conception of epistemic success, dominant in much of twentieth-century epistemology (cf. Kvanvig, 2003; Greco, 2014).

This picture becomes increasingly strained once attention is paid to the actual representational practices employed across different cognitive domains. Scientific inquiry relies extensively on models, idealizations, and abstractions that are known not to be literally accurate descriptions of reality. These representations are nevertheless indispensable for explanation, prediction, and reasoning about complex systems. Their epistemic value does not lie simply in stating true propositions, but in enabling agents to grasp patterns, dependencies, and relations that would otherwise remain opaque through forms of representation that invite interpretation rather than mere belief formation. As has been widely noted in recent philosophy of science, the intelligibility and usability of models often matter more for understanding than their strict descriptive accuracy. Accounts of scientific understanding have therefore increasingly emphasized intelligibility, use, and explanatory traction over representational fidelity (de Regt, 2017; Elgin, 2017).

Parallel considerations arise in the case of art and fiction. Narrative and fictional representations are not ordinarily engaged with as sources of factual information, yet they are widely taken to afford insight into human experience, social relations, and forms of agency. What such representations provide is typically not new knowledge about how the world actually is, but an enhanced capacity to make sense of experiences, perspectives, and possibilities. Their cognitive contribution consists less in the transmission of information than in the reorganization of meaning and attention, a kind of epistemic gain that resists straightforward reduction to propositional content. This form of understanding is often holistic, context-sensitive, and resistant to paraphrasing, a feature frequently highlighted in discussions of understanding as distinct from knowledge (Grimm, 2006; Hills, 2016).

Despite these similarities, philosophical discussions of understanding in science and in art have often developed independently. Accounts of scientific understanding tend to focus on explanation, modeling, and idealization, while discussions of fiction emphasize imagination, narrative structure, and experiential insight. As a result, the deeper commonalities between these practices, particularly their reliance on shared representational conventions, interpretive norms, and criteria of appropriate use, have received comparatively little systematic attention. In particular, the role of norm-governed engagement with representations as a common source of cognitive gain across domains remains underexplored. While recent work has begun to address understanding as a practice-sensitive and gradable epistemic achievement (Baumberger et al., 2017), a unified account spanning science, fiction, and technologically mediated representation is still lacking.

The aim of this paper is to articulate a unified account of understanding that applies across these domains. The central thesis is that many significant forms of cognitive gain are better characterized as instances of understanding rather than as instances of knowledge. Understanding, on the account defended here, is not primarily a matter of possessing accurate beliefs, but of being able to competently interpret,

navigate, and reason within representational frameworks that structure inquiry, interpretation, and reasoning. These frameworks guide attention, constrain inference, and enable agents to explore possibilities, including counterfactual and non-actual scenarios, in ways that are governed by normative expectations rather than by truth alone. This conception aligns with practice-oriented and non-factivist approaches to understanding, according to which epistemic success is grounded in use, intelligibility, and inferential competence rather than belief possession alone (Dellsén, 2017; de Regt, 2017).

From this perspective, fictional and idealized elements are not epistemic shortcomings to be eliminated, but constitutive features of many understanding-generating practices. In science, models abstract and distort in order to render complex systems intelligible. In art and fiction, narrative representations organize experience in ways that make salient patterns of meaning, motivation, and agency. In both cases, understanding depends on the successful uptake of representations within practices that specify how they are to be interpreted, evaluated, and used. Accordingly, epistemic evaluation shifts from assessing representations in isolation to assessing the practices and norms that govern their employment. This shift resonates with accounts that defend the epistemic legitimacy of idealization and fiction when embedded in appropriate normative frameworks (Cartwright, 1983; Elgin, 2017).

The paper proceeds as follows. Section 2 clarifies the conceptual distinction between knowledge and understanding. Section 3 examines the epistemic role of fictional and idealized models in scientific practice. Section 4 turns to art and fiction, analyzing their contribution to cognitive gain. Section 5 develops a general account of understanding as structured and norm-governed cognitive engagement, and Section 6 illustrates this framework through cases drawn from science, fiction, and AI-mediated interpretation. The conclusion reflects on the implications of this account for contemporary epistemology and the theory of representation.

2. Understanding and knowledge: distinct epistemic achievements

Philosophical discussions of cognition have traditionally treated knowledge as the central epistemic achievement. Within this framework, to know is typically understood as possessing true beliefs supported by adequate justification, and epistemic evaluation is largely oriented toward questions of truth, evidence, and reliability. The dominant questions concern how beliefs are formed, whether they are true, and whether they are properly justified or reliably produced. Understanding, when acknowledged at all, has often been treated as a derivative notion, a pragmatic or pedagogical supplement to knowledge rather than a theoretically autonomous epistemic category. This orientation is characteristic of mainstream epistemology, where understanding is frequently subsumed under knowledge or treated as epistemically secondary, lacking independent criteria of success (Kvanvig, 2003; Grimm, 2006).

This traditional orientation reflects an implicit assumption: that the primary aim of cognition is the accurate representation of the world, and that epistemic success consists in matching beliefs to facts. From this perspective, understanding is frequently

conceived as a mere byproduct of knowledge acquisition, a psychological state accompanying the possession of many true beliefs, or a heuristic label for especially well-organized knowledge. Such views leave little conceptual space for understanding as an independent epistemic achievement with its own normative profile. As a result, the epistemic significance of practices that do not primarily aim at literal truth has often been underestimated or misunderstood.

Over the last decades, however, this hierarchy has been increasingly questioned. A growing body of philosophical work argues that understanding constitutes a distinct epistemic achievement, one that cannot be fully reduced to the possession of true propositions. While knowledge is paradigmatically factive and belief-centered, understanding appears to involve grasping relations, patterns, and dependencies that organize information into an intelligible whole (Grimm, 2010; de Regt, 2017). Crucially, such grasp may persist even in the absence of complete, precise, or fully accurate information, so long as the relevant relations can be appropriately interpreted and navigated. This shift has been especially prominent in philosophy of science, where attention to scientific practice has revealed forms of epistemic success not well captured by traditional knowledge-centered models (Dellsén, 2017).

One of the central motivations for this shift comes from reflection on ordinary epistemic practices. It is widely acknowledged that an agent may possess a large body of correct information about a domain and yet fail to understand it. Someone might know many isolated facts about a scientific theory, a historical period, or a social phenomenon while lacking any sense of how those facts fit together, why they matter, or what explains them. Conversely, agents are often credited with understanding even when their representations are simplified, schematic, or known to be inaccurate in certain respects. These everyday judgments suggest that understanding involves more than the accumulation of truths; it involves the ability to make sense of information within a coherent representational perspective. This asymmetry between knowledge and understanding challenges the assumption that epistemic value is exhausted by truth possession (Grimm, 2006).

A prominent line of argument emphasizes that understanding requires insight into how and why certain facts hold, not merely knowledge that they do. Understanding is associated with the ability to explain, to answer “why” and “how” questions, and to see how different elements of a domain depend on one another. This explanatory dimension distinguishes understanding from knowledge conceived narrowly as justified true belief. Knowing that a phenomenon occurs does not by itself amount to understanding why it occurs or how it relates to other phenomena (Lipton, 2004; Grimm, 2010).

This point becomes especially salient when considering the role of integration. Understanding is characteristically holistic: it involves seeing how pieces of information form a coherent structure. An agent who understands a domain can situate particular facts within a broader framework, recognize which factors are explanatorily central, and distinguish relevant from irrelevant details. Knowledge, by contrast, may remain fragmented. One can know many truths without possessing the unifying perspective that understanding provides, or without being able to deploy those truths

within a meaningful inferential pattern. This contrast supports the idea that understanding has an essentially structural dimension that is absent from many accounts of knowledge (Kvanvig, 2003).

These considerations have led many philosophers to treat understanding as a higher-order epistemic achievement. Rather than consisting in the possession of individual beliefs, understanding involves the organization, coordination, and application of information within a structured representational framework. This framework enables agents to reason effectively, to make sense of new information, and to extend their grasp of the domain beyond what is explicitly known, including by guiding interpretation in novel or non-standard cases (Baumberger, Beisbart, and Brun, 2017).

The distinction between knowledge and understanding has been particularly influential in philosophy of science. Scientific practice provides numerous examples in which understanding is achieved through representations that are known to be idealized or even false. Scientific models often deliberately simplify complex systems, omit causal factors, or introduce fictional elements in order to make patterns and dependencies salient. These representations are not typically regarded as epistemic failures. On the contrary, they are central to explanation, prediction, and theoretical reasoning. Their epistemic success is commonly attributed to their capacity to render phenomena intelligible rather than to their strict descriptive accuracy (Cartwright, 1983; de Regt and Gijsbers, 2017).

Scientific understanding, in this context, is commonly associated with intelligibility. A theory or model is said to promote understanding when it renders a phenomenon intelligible to its users, enabling them to reason about it, to answer explanatory questions, and to explore counterfactual scenarios. Importantly, intelligibility is not a purely semantic property of representations, nor does it require strict truth at every point. It depends on how representations function within a practice and on the skills agents bring to their use, including their capacity to interpret representational elements in accordance with shared norms (de Regt, 2015; 2017).

From this perspective, understanding involves epistemic abilities rather than merely epistemic states. To understand a phenomenon is to be able to use a theory or model competently: to manipulate it, to draw inferences from it, to recognize its scope and limitations, and to apply it to novel situations. These abilities may be exercised even when the underlying representation includes idealizations or known inaccuracies. What matters is not perfect correspondence with reality, but the capacity of the representational system to support reliable and illuminating reasoning through appropriate modes of use (Dellsén, 2017).

Closely related to this emphasis on intelligibility is the idea that understanding has a counterfactual dimension. An agent who understands a phenomenon is typically able to answer questions about how it would behave under different conditions, what would happen if certain factors were altered, or which aspects of the system are responsible for particular outcomes. Such counterfactual competence goes beyond knowing what actually happens; it involves grasping the structure of dependencies that govern the phenomenon as they are represented within a model or framework, a point

emphasized in accounts that link understanding to the ability to explore counterfactual dependencies through models (Kuorikoski, 2011).

This conception of understanding helps explain why it is compatible with epistemic imperfection. Models that contain false assumptions can still support understanding if they accurately capture relevant dependencies and allow users to reason effectively about the system. In this sense, understanding tolerates a degree of representational inaccuracy that knowledge, strictly construed, does not. This tolerance is not a weakness, but a reflection of the different epistemic roles played by understanding and knowledge (Elgin, 2017).

The distinction between knowledge and understanding also has normative implications. Knowledge is typically evaluated in terms of truth and justification: a belief either meets the relevant standards or it does not. Understanding, by contrast, admits of degrees and is sensitive to context. One may understand a phenomenon better or worse, depending on the richness of the representational framework, the range of counterfactuals one can handle, and the purposes for which the understanding is deployed. This gradability further distinguishes understanding from knowledge (Baumberger, 2019).

Seen in this light, understanding occupies a distinct epistemic space between mere belief and full theoretical knowledge. It is structured, relational, and practice-oriented. It depends on mastery of representational frameworks and inferential norms rather than on the passive possession of truths. Understanding is achieved through engagement: through using models, narratives, diagrams, or other representational devices to make sense of complex domains in ways that are governed by norms of interpretation rather than by truth alone.

This reconceptualization has important consequences for how fictional and idealized representations are evaluated. If understanding does not require strict factual accuracy at every representational level, then the presence of fictional elements in epistemic practices need not be regarded as epistemically suspect. On the contrary, such elements may be essential for highlighting salient relations, constraining interpretation, and supporting exploratory reasoning. Fiction and idealization can function as cognitive tools rather than as epistemic obstacles.

Moreover, this account helps explain why understanding is often robust under revision. Scientific theories change, models are replaced, and representations are refined, yet the understanding they afford may persist or even deepen. This persistence suggests that understanding is not tied to any particular set of propositions, but to the broader representational practices through which agents engage with a domain.

The distinction between knowledge and understanding thus sets the stage for the analyses that follow. If understanding is a distinct epistemic achievement grounded in structured engagement with representations, then it becomes possible to reassess the epistemic roles of scientific models and works of fiction. The next section examines scientific modeling as a paradigmatic case in which fictional and idealized representations contribute centrally to understanding.

3. Scientific models and epistemic fiction

Scientific practice relies extensively on models that simplify, idealize, or deliberately distort aspects of reality. From frictionless planes in classical mechanics to ideal gases in thermodynamics and highly simplified agents in economics, many of the central representational tools of science are known to be literally false. They do not aim to describe the world in full detail, nor do they aspire to capture every causal factor at work in the phenomena they represent. Yet these models are not treated as epistemic failures. On the contrary, they are widely regarded as indispensable for explanation, prediction, and, most importantly, scientific understanding (Cartwright, 1983; de Regt, 2017).

This situation poses an apparent puzzle. If epistemic success were exhausted by the accurate representation of facts, the pervasive reliance on idealized and inaccurate models would be difficult to justify. Scientific inquiry would seem systematically compromised by its own methods. The persistence and centrality of modeling practices therefore motivate a reassessment of the epistemic standards by which scientific representations are evaluated, shifting attention from accuracy alone to the roles representations play within inquiry (Mizrahi, 2012; Dellsén, 2017).

Rather than viewing models as incomplete descriptions awaiting correction, contemporary philosophy of science increasingly emphasizes their functional role within scientific inquiry. Models are not primarily mirrors of reality, but representational instruments designed to make complex systems intelligible. They enable scientists to isolate dependencies, explore counterfactual scenarios, and identify explanatory patterns that would remain opaque if all details were retained. Their epistemic value lies less in their literal truth than in what they enable users to do with them, namely, to reason, explain, and explore within a structured representational space, including the exploration of counterfactual dependencies and the generation of intelligibility despite idealization (Kuorikoski, 2011; de Regt and Gijsbers, 2017).

On this view, scientific models often function as disciplined or constrained fictions. They introduce simplified systems, ideal conditions, or counterfactual assumptions that are known not to obtain in the world. These fictional elements are not arbitrary inventions. They are governed by theoretical commitments, empirical constraints, and methodological norms that regulate how the model is constructed, interpreted, and applied. The departure from literal accuracy is deliberate and controlled, serving specific cognitive purposes by guiding interpretation rather than by asserting facts (Elgin, 2017).

The epistemic contribution of such models is best understood in terms of intelligibility. A model promotes understanding when it renders a phenomenon cognitively tractable for its users. This involves more than the communication of isolated truths. It requires that the model support coherent reasoning, explanation, and exploration. A model is intelligible insofar as scientists can use it to answer "how" and "why" questions, to anticipate how the system would behave under hypothetical changes, and to situate particular observations within a broader explanatory framework that organizes significance and relevance (de Regt, 2015; de Regt, 2017).

Importantly, intelligibility is not a purely representational property. It depends on the interaction between the model and the skills of its users. To understand a phenomenon through a model is to be able to manipulate it, to recognize the significance of its assumptions, and to draw appropriate inferences from it. These abilities may be exercised even when the model includes assumptions that are strictly false. What matters epistemically is not perfect correspondence with reality, but the model's capacity to guide competent reasoning within accepted norms of interpretation and use (Grimm, 2010; Dellsén, 2017).

This perspective helps explain why idealization is often epistemically productive. By omitting or distorting certain features of a system, a model can make salient relations that would otherwise be obscured. Too much detail can hinder understanding by masking patterns and dependencies. Simplification, in this context, is not a loss but a gain, insofar as it enables the isolation of explanatorily relevant dependencies and the stabilization of interpretive use across contexts (Cartwright, 1983; Kuorikoski, 2011).

From this standpoint, distortion is not an epistemic defect but a methodological resource. A model may misrepresent the exact behavior of a system while still accurately capturing how key variables interact. It may fail to describe what actually happens in every circumstance while nonetheless illuminating why certain outcomes occur or how they would change under different conditions. The understanding afforded by the model resides in these structural insights, not in the literal truth of each assumption taken in isolation (Elgin, 2017).

This account also sheds light on the resilience of scientific understanding under theoretical revision. As empirical knowledge advances, models are refined, replaced, or abandoned. Yet the understanding they provide often persists or even deepens. This persistence suggests that the cognitive gain associated with a model is not exhausted by the particular propositions it encodes. Instead, it is grounded in the inferential practices, explanatory perspectives, and counterfactual reasoning the model makes possible across successive representational frameworks (Dellsén, 2016).

The role of fiction in scientific modeling further highlights the importance of conventions and shared interpretive norms. Scientists must coordinate on how a model is to be used, which idealizations are to be taken seriously, and which aspects of the representation are to be bracketed. A frictionless plane, for example, is not taken to deny the existence of friction, but to suspend it for the sake of isolating other dependencies. Understanding arises not from the model alone, but from a socially regulated practice that governs how the fictional scenario is connected to real-world phenomena through established interpretive rules (de Regt and Gijsbers, 2017).

This underscores the normative and practice-dependent character of scientific understanding. Understanding is not a private mental state reducible to belief possession. It is an achievement embedded in communal practices of modeling, interpretation, and explanation. The epistemic success of a model depends on how it is used within these practices, not merely on its representational accuracy or its correspondence to isolated facts (Baumberger, Beisbart, and Brun, 2017).

Seen in this light, scientific models exemplify a broader epistemic pattern. Understanding is achieved through structured engagement with representational systems that are neither fully literal nor purely subjective. Fictional elements are harnessed in a controlled and norm-governed manner to facilitate insight, orientation, and explanatory control. Far from undermining rational inquiry, such elements are integral to the ways in which science makes the world intelligible.

As the next section argues, this epistemic pattern is not unique to science. It also characterizes the cognitive contributions of art and fiction, where understanding likewise emerges from imaginative engagement guided by shared conventions rather than from the transmission of literal truths.

4. Art, fiction, and cognitive Gain

While scientific models provide a paradigmatic case of epistemically productive idealization, similar patterns of cognitive gain can be observed in the domain of art and fiction. Unlike scientific representations, fictional narratives and artworks are not typically evaluated in terms of truth or empirical accuracy. They are openly non-factive and are engaged with under conventions that suspend literal belief. Yet despite this, works of fiction are widely credited with yielding genuine insight into human experience, social structures, moral conflict, and forms of agency (Elgin, 2017; Currie, 2020).

This apparent tension has long posed a challenge for epistemology. If epistemic value is tied essentially to truth, then the cognitive significance commonly attributed to fiction would appear misguided or merely metaphorical. However, a growing body of work in aesthetics and epistemology argues that fiction can contribute to understanding without functioning as a source of propositional knowledge. The relevant epistemic achievement is not the acquisition of new facts, but the reorganization of attention, perspective, and interpretive frameworks through which agents make sense of the world (Gaut, 2003; Elgin, 2017).

Fictional narratives operate by constructing imagined scenarios, characters, and events that are not intended to describe reality as it is. Instead, they present structured representations that invite readers or viewers to explore possibilities, counterfactuals, and patterns of significance. What is epistemically valuable in such engagement is not belief in the content of the fiction, but the disciplined imaginative activity it supports. Through narrative structure, point of view, and symbolic articulation, fiction enables agents to grasp relationships, motivations, and constraints that may be difficult to articulate or apprehend through literal description alone (Currie, 2010; Hills, 2016).

In this respect, fiction shares important functional similarities with scientific models. Both employ departures from literal accuracy in order to enhance intelligibility. Just as idealized models isolate causal dependencies by bracketing irrelevant factors, fictional narratives isolate experiential, moral, or social structures by abstracting away from empirical contingency. The resulting representations are not evaluated primarily

in terms of truth, but in terms of how effectively they support interpretation, reflection, and understanding within established practices of use.

The epistemic contribution of fiction is therefore best understood as perspectival and structural rather than informational. Works of literature, film, or other narrative media can deepen understanding by reorganizing how agents perceive and interpret familiar phenomena. They can make salient patterns of action, forms of vulnerability, or normative tensions that are otherwise overlooked, not by stating them explicitly, but by embedding them within coherent narrative structures. This kind of understanding is holistic and context-sensitive, resisting reduction to a set of discrete propositions (Elgin, 2017; Currie, 2020).

Crucially, the cognitive gains afforded by fiction depend on shared interpretive conventions. Readers do not approach fictional narratives as assertions about the actual world, but as invitations to imagine under specific constraints. These conventions regulate what kinds of inferences are appropriate, which features of the narrative are salient, and how fictional content may legitimately inform real-world reflection. Without such norms, imaginative engagement would collapse into arbitrary fantasy, and fiction would lose its capacity to generate understanding.

This norm-governed character explains why not all fictions are equally epistemically valuable. Just as poorly constructed scientific models fail to support understanding, incoherent or norm-violating narratives fail to yield meaningful insight. Epistemic evaluation in the aesthetic domain therefore concerns the adequacy of representational structure, coherence, and interpretive affordances rather than factual accuracy. Understanding arises when imaginative engagement is guided by stable conventions that support disciplined exploration rather than unconstrained invention (Gaut, 2003; Elgin, 2017).

The role of counterfactual reasoning is particularly salient in fiction. Narrative representations invite agents to consider how individuals might act under different circumstances, how social structures shape agency, or how values conflict in non-actual scenarios. This counterfactual exploration expands the space of intelligibility by allowing agents to grasp dependencies and possibilities without requiring empirical instantiation. As in scientific modeling, the epistemic payoff lies in the ability to navigate these possibilities in a principled way, not in their literal truth.

Understanding through fiction is also compatible with epistemic imperfection. Fictional narratives may exaggerate, simplify, or stylize aspects of human life. Such distortions do not undermine their epistemic value so long as they serve to illuminate structurally relevant relations. Indeed, exaggeration and stylization often play a role analogous to idealization in science, enabling insight by foregrounding what matters most for interpretive purposes (Elgin, 2017).

These considerations support a broader claim about understanding as an epistemic achievement. Understanding does not require that representations be true, but that they be usable within norm-governed practices that support coherent interpretation, counterfactual reasoning, and reflective application. Art and fiction exemplify

this mode of epistemic engagement particularly clearly, revealing how cognitive gain can be achieved through imaginative structures rather than through belief formation.

By situating fiction alongside scientific modeling, this analysis challenges the sharp epistemic divide often drawn between science and art. Both domains employ representational fictions to generate understanding, differing primarily in their aims, constraints, and modes of evaluation rather than in their epistemic legitimacy. Recognizing this continuity prepares the ground for a general account of understanding that treats structured engagement with representations as its core feature.

The next section develops this account explicitly, articulating understanding as a form of structured and norm-governed cognitive engagement that applies across scientific, artistic, and technologically mediated contexts alike.

5. Understanding as structured engagement

The preceding sections suggest a common epistemic pattern underlying both scientific modeling and artistic fiction. In neither case is cognitive success best explained in terms of the accumulation of true propositions. Instead, understanding emerges from a form of engagement with representational systems that guide attention, structure interpretation, and enable meaningful inference. This approach aligns with recent work that treats understanding as a distinct epistemic achievement irreducible to belief or information possession (Grimm, 2014; Elgin, 2017). This section develops a general account of understanding along these lines, characterizing it as a form of structured cognitive engagement rather than as a mental state defined by belief or truth alone.

On this view, understanding is an epistemic achievement that involves the active coordination of multiple cognitive elements: representational artifacts, interpretive conventions, background assumptions, and inferential practices. To understand a phenomenon is not merely to know certain facts about it, but to be able to situate those facts within an organized framework that renders their relations, relevance, and implications intelligible. This emphasis on organization and relational grasp reflects objectual and holistic accounts of understanding, according to which understanding targets structured domains rather than isolated propositions (Kvanvig, 2003; Khalifa, 2017). Such frameworks are typically provided by models, narratives, diagrams, or symbolic systems that impose structure on domains that would otherwise remain cognitively opaque by making salient what counts as relevant, explanatory, or significant.

This conception departs from accounts that locate understanding primarily in an internal psychological state or in the possession of particular beliefs. While mental attitudes undoubtedly play a role, understanding cannot be reduced to what an individual believes in isolation. Rather, it is essentially relational and practice-dependent. It depends on the agent's ability to engage competently with external representational systems and to navigate the norms governing their use. In this respect, understanding is better characterized in terms of epistemic abilities and competencies than in terms

of belief states (Grimm, 2012; de Regt, 2017). Understanding is thus as much a matter of epistemic participation as of mental endorsement, involving mastery of ways of using representations rather than mere assent to their content.

A central feature of structured engagement is its reliance on conventions. Representational systems function epistemically only insofar as users share norms governing how those systems are to be interpreted and employed. In scientific modeling, such conventions determine which elements of a model are idealized, which are intended to track features of the target system, and how inferences drawn within the model may legitimately be projected onto reality. In art and fiction, analogous conventions regulate imaginative uptake, the suspension of literal belief, and the relevance of fictional content to real-world concerns by specifying how meaning is to be taken rather than what is literally asserted. This norm-governed dimension of representation has been emphasized in accounts of scientific understanding centered on intelligibility and use (de Regt and Dieks, 2005; de Regt, 2017).

Importantly, these conventions do not merely constrain interpretation; they actively enable understanding. By stabilizing expectations and coordinating inferential practices, conventions allow fictional or idealized representations to support coherent reasoning, comparison, and reflection. Without such normative structure, imaginative engagement would collapse into arbitrary fantasy, and models would lose their epistemic grip on the world. Understanding thus depends less on representational accuracy at the level of individual elements than on the stability and appropriateness of the interpretive framework within which those elements function (Elgin, 2017). It is within such frameworks that representations acquire epistemic significance.

Another defining feature of understanding as structured engagement is its counterfactual dimension. Both scientific models and fictional narratives invite users to explore possibilities, alternatives, and variations that need not be actual. The capacity to reason counterfactually, to ask how a system would behave under different conditions, or how an agent might respond in a different context, is central to explanatory and interpretive insight. Counterfactual competence is widely regarded as a hallmark of understanding rather than mere knowledge (Woodward, 2003; Khalifa, 2017). Fictional elements facilitate this process by loosening the constraints of actuality while preserving structurally relevant relations.

This account also helps explain why understanding is often resilient in the face of error. A model may incorporate false assumptions, and a fictional narrative may rely on implausible or even impossible scenarios, yet both can still generate genuine understanding if they succeed in organizing experience in a coherent and explanatorily productive way. What matters is not the literal truth of each representational component, but the overall capacity of the system to support stable interpretation, counterfactual reasoning, and guided application. This tolerance for idealization and distortion has been widely recognized in contemporary accounts of scientific understanding (de Regt, 2017; Elgin, 2017).

Understanding, therefore, should not be conceived as a static epistemic state, but as a dynamic and ongoing practice. It involves skills, dispositions, and interpretive

competencies exercised through sustained engagement with representational artifacts. This practical and normative dimension distinguishes understanding from knowledge conceived narrowly as justified true belief. While knowledge can, at least in principle, be possessed passively, understanding requires active participation in a structured cognitive activity governed by shared norms and standards of use that regulate how representations are taken to mean and how they may be extended to new cases.

By framing understanding in this way, the apparent epistemic tension between fiction and cognition dissolves. Fictional and idealized representations are not obstacles to understanding, but integral components of many understanding-generating practices. When embedded within stable conventions and employed for disciplined exploration, they enable forms of insight that would be inaccessible through strictly literal or purely propositional means alone.

The final section of the paper applies this account to a set of illustrative cases drawn from science, fiction, and contemporary representational technologies. These cases clarify how structured engagement operates across different domains and show how a unified conception of understanding can illuminate the epistemic roles of models, narratives, and mediated representations alike.

6. Illustrative cases: science, fiction, and AI-mediated interpretation

The account of understanding developed in the previous sections gains depth and precision when examined through concrete epistemic practices. This section considers three domains in which structured engagement plays a central role in generating understanding: scientific modeling, literary fiction, and AI-mediated interpretation. Although these practices differ in aims, methods, and institutional settings, they share a reliance on fictional, idealized, or non-literal representations governed by stable conventions of use. Recent work in philosophy of science and aesthetics has increasingly emphasized these shared structural features (de Regt, 2017; Elgin, 2017). Examining these cases helps clarify how understanding arises through engagement with representational systems rather than through the mere possession of true beliefs or the passive reception of information.

6.1. Scientific models and idealization

Scientific models routinely employ idealizations and known falsehoods. Frictionless planes, perfectly rational agents, isolated systems, and homogeneous populations are familiar examples across physics, economics, and biology. From the standpoint of a strictly truth-centered epistemology, such representations appear epistemically deficient, since they misdescribe the systems they are meant to represent. Yet in scientific practice, these models are treated as central vehicles of understanding rather than as provisional approximations awaiting elimination.

On the present account, the epistemic role of such models is best explained in terms of structured engagement. Models generate understanding by enabling scientists to explore dependency relations, isolate explanatorily relevant factors, and reason

counterfactually about how systems would behave under varying conditions. This functional conception of models is widely defended in contemporary philosophy of science, particularly in accounts that emphasize intelligibility over literal accuracy (de Regt and Dieks, 2005; Weisberg, 2013). What is gained is not primarily a set of propositions believed to be true, but a practical grasp of how to navigate a space of possibilities in a disciplined and informative manner according to shared inferential standards.

Importantly, this understanding can persist even when the idealizations involved are explicitly acknowledged as false. A physicist does not believe that friction is absent, nor does an economist believe that agents are perfectly rational. What matters epistemically is mastery of the inferential practices associated with the model: knowing when it applies, how its assumptions constrain its use, and how its results can be related back to empirical systems. Idealization thus functions as an epistemic strategy rather than as a representational defect (Elgin, 2017; Frigg and Nguyen, 2020).

This case illustrates a key feature of understanding as structured engagement: epistemic success depends less on representational fidelity at the level of individual assumptions than on the stability and productivity of the interpretive framework within which reasoning takes place. Scientific understanding emerges from participation in a norm-governed practice of model use, not from assent to the literal content of the model itself taken in isolation.

6.2. Literary fiction and experiential understanding

A parallel structure can be observed in the case of literary fiction. Readers typically do not approach novels, plays, or films with the expectation of acquiring factual knowledge about the world. Nevertheless, engagement with fiction is widely regarded as cognitively valuable, yielding understanding of human motivation, social interaction, moral conflict, and emotional life.

This cognitive gain is not reducible to the extraction of general truths from narrative content. Rather, fiction affords structured experiential access to possibilities that may be rare, inaccessible, or ethically unavailable in real life. Philosophers of art and literature have emphasized that such understanding arises through imaginative engagement governed by narrative and genre conventions (Currie, 1995; Gaut, 2003). By imaginatively inhabiting fictional scenarios, readers explore perspectives and situations in a controlled environment shaped by narrative conventions. These conventions guide attention, regulate interpretation, and constrain imaginative engagement in ways that support coherent reflection and meaningful comparison with real-world experience.

From the standpoint of structured engagement, fiction functions as an epistemic laboratory. It allows readers to test responses, evaluate choices, and recognize patterns of agency and vulnerability without committing to the belief that the represented events actually occurred. The understanding gained is holistic and non-propositional, involving sensitivity to salience, context, and relevance rather than the endorsement of explicit generalizations (Elgin, 2017; Currie, 2020).

This perspective helps explain why fictional understanding is often resistant to paraphrasing. What is grasped through narrative engagement cannot always be captured by a list of statements, even when such statements are true. The epistemic value of fiction lies in its capacity to organize experience and cultivate interpretive skill, not in its contribution to the stock of factual knowledge considered independently of context.

6.3. AI systems and interpretive mediation

Recent developments in artificial intelligence offer a contemporary extension of this framework. Large language models and other AI systems do not possess beliefs, intentions, or understanding in any robust philosophical sense. Nonetheless, they increasingly mediate human understanding by generating representations with which users engage interpretively through linguistic interaction.

AI-generated outputs often involve simplifications, generalizations, and occasional errors. Evaluated purely in terms of truth or reliability, such systems may appear epistemically fragile. Yet in practice, they can support understanding when embedded within appropriate conventions of use. Recent philosophical work on scientific representation and cognitive tools emphasizes that epistemic value often lies in how representational artifacts are used rather than in their intrinsic accuracy (de Regt, 2017; Frigg and Nguyen, 2020).

On the present account, the epistemic contribution of AI systems resides not in the representations they produce in isolation, but in the structured interaction they enable. When users engage reflectively with AI-generated content, testing it against background knowledge and situating it within broader interpretive frameworks, cognitive gain can occur despite imperfections in output. Understanding arises from the human practice of use, not from the machine's internal processes or its capacity to store or retrieve information.

This case reinforces a central claim of the paper: understanding is not located solely in representational accuracy or informational content, but in norm-governed practices of engagement. Even representations that are partially fictional, idealized, or unreliable can contribute to understanding when used within stable and reflective interpretive conventions that regulate how they are taken to mean and how they are integrated into ongoing inquiry.

7. Conclusion: understanding as structured cognitive engagement

This paper has argued that understanding constitutes a distinct epistemic achievement that cannot be reduced to the possession of true propositions or justified belief. By examining epistemic practices across science, fiction, and contemporary forms of technological mediation, it has shown that understanding often arises through engagement with representations that are idealized, fictional, or partially inaccurate, yet systematically constrained by norms of use and interpretation.

The framework proposed here characterizes understanding as structured cognitive engagement. On this view, epistemic success does not hinge on strict factual correspondence at the level of individual representations, but on the presence of normative constraints, interpretive conventions, and inferential affordances that guide cognitive interaction. Scientific models, literary narratives, and AI-generated representations exemplify this structure in different ways. Each relies on controlled departures from literal truth to facilitate intelligibility, orientation, and insight within shared practices of sense-making, a point increasingly emphasized in contemporary accounts of understanding and intelligibility (de Regt, 2017; Elgin, 2017).

This account helps explain why fictional and idealized elements can play a legitimate epistemic role without undermining rational inquiry. In science, idealizations enable the isolation of causal patterns and explanatory relations that would otherwise remain obscured. In fiction, narrative constructions afford experiential access to psychological, moral, and social possibilities that resist purely propositional articulation. In AI-mediated contexts, simplified and probabilistic outputs support exploration, comparison, and conceptual clarification when embedded within appropriate interpretive practices. In all cases, understanding emerges from use and engagement rather than from representational content considered in isolation or assessed solely by its truth conditions.

By emphasizing engagement over belief, the framework also clarifies the relationship between understanding and knowledge. Knowledge remains indispensable where accuracy, justification, and reliability are the primary epistemic aims. Understanding, however, operates with a different normative profile. It tolerates non-literal representations, counterfactual reasoning, and controlled distortions, provided that these contribute to coherent cognitive orientation and explanatory competence. In this sense, understanding is compatible with epistemic humility: one can understand a phenomenon without claiming exhaustive or final knowledge of it, or full propositional mastery of its details.

The analysis contributes to ongoing debates in philosophy of science, aesthetics, and epistemology by offering a unified account of epistemic practices that are often treated in isolation. It suggests that the apparent divide between science and art reflects differences in representational aims rather than differences in epistemic value. Both domains employ fictions and conventions to cultivate understanding, and both can be evaluated according to how effectively they structure cognitive engagement and guide interpretation.

Finally, the proposed framework has implications for contemporary epistemology in an age of increasing technological mediation. As AI systems become integral to scientific research, education, and public discourse, assessing their epistemic role requires moving beyond narrowly truth-centric models. Understanding how such systems can enhance, rather than obscure, human cognition depends on recognizing the normative practices that govern their use and the interpretive competencies of their users.

Understanding beyond knowledge, the paper concludes, is not an epistemic deficiency but a fundamental mode of human cognition: one through which agents make sense of the world by engaging with representations that are intelligible, disciplined, and normatively constrained, even when they are not strictly true.

References

Baumberger, C. (2019). Explicating Objectual Understanding: Taking Degrees Seriously. *Journal for General Philosophy of Science*, 50, 367–388.

Baumberger, C., C. Beisbart, and G. Brun. (2017). What Is Understanding? An Overview of Recent Debates in Epistemology and Philosophy of Science. In *Explaining Understanding: New Perspectives from Epistemology and Philosophy of Science*, edited by Stephen R. Grimm, Christoph Baumberger, and Sabine Ammon. New York: Routledge.

Cartwright, N. (1983). *How the Laws of Physics Lie*. Oxford: Oxford University Press.

Currie, G. (1995). *Image and Mind: Film, Philosophy and Cognitive Science*. Cambridge: Cambridge University Press.

Currie, G. (2010). *Narratives and Narrators: A Philosophy of Stories*. Oxford: Oxford University Press.

Currie, G. (2020). *Imagining and Knowing: The Shape of Fiction*. Oxford: Oxford University Press.

Dellsén, F. (2016). Scientific Progress: Knowledge versus Understanding. *Studies in History and Philosophy of Science*, 56, 72–83.

Dellsén, F. (2017). Understanding without Justification or Belief. *Ratio*, 30(3), 239–254.

de Regt, H. W. (2014). Visualization as a Tool for Understanding. *Perspectives on Science*, 22(3), 377–396.

de Regt, H. W. (2015). Scientific Understanding: Truth or Dare? *Synthese*, 192, 3781–3797.

de Regt, H. W. (2017). *Understanding Scientific Understanding*. Oxford: Oxford University Press.

de Regt, H. W. and D. Dieks. (2005). A Contextual Approach to Scientific Understanding. *Synthese*, 144, 137–170.

de Regt, H. W., and Victor Gijsbers. 2017. "How False Theories Can Yield Genuine Understanding." *In Explaining Understanding: New Perspectives from Epistemology and Philosophy of Science*, edited by Stephen R. Grimm, Christoph Baumberger, and Sabine Ammon. New York: Routledge.

Elgin, C. Z. (2017). *True Enough*. Cambridge, MA: MIT Press.

Frigg, R. and J. Nguyen. (2020). *Modelling Nature: An Opinionated Introduction to Scientific Representation*. Cham: Springer.

Gaut, B. (2003). Art and Knowledge. In *The Oxford Handbook of Aesthetics*, edited by Jerrold Levinson. Oxford: Oxford University Press.

Grimm, S. R. (2006). Is Understanding a Species of Knowledge? *British Journal for the Philosophy of Science*, 57(3), 515–535.

Grimm, S. R. (2010). The Goal of Explanation. *Studies in History and Philosophy of Science*, 41 (4), 337–344.

Grimm, S. R. (2014). Understanding as Knowledge of Causes. In *Virtue Epistemology Naturalized*, edited by Abrol Fairweather, 329–345. Cham: Springer.

Grimm, S. R. (2017). Understanding and Transparency. In *Explaining Understanding: New Perspectives from Epistemology and Philosophy of Science*, edited by Stephen R. Grimm, Christoph Baumberger, and Sabine Ammon. New York: Routledge.

Hills, A. (2016). *Understanding Why*. Oxford: Oxford University Press.

Khalifa, K. (2017). *Understanding, Explanation, and Scientific Knowledge*. Cambridge: Cambridge University Press.

Kuorikoski, J. (2011). Simulation and the Sense of Understanding. In *Models, Simulations, and Representations*, edited by Paul Humphreys and Cyrille Imbert. New York: Routledge.

Kvanvig, J. L. (2003). *The Value of Knowledge and the Pursuit of Understanding*. Cambridge: Cambridge University Press.

Lipton, P. (2004). *Inference to the Best Explanation*. 2nd ed. London: Routledge.

Mizrahi, M. (2012). Idealizations and Scientific Understanding. *Philosophical Studies*, 160(2), 237–252.

Weisberg, M. (2013). *Simulation and Similarity: Using Models to Understand the World*. Oxford: Oxford University Press.

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.