

---

# The Universal Landscape of Human Reasoning

---

Qiguang Chen<sup>1\*</sup> Jinhao Liu<sup>1\*</sup> Libo Qin<sup>2†</sup> Yimeng Zhang<sup>3</sup> Yihao Liang<sup>4</sup>  
 Shangxu Ren<sup>1</sup> Chengyu Luan<sup>1</sup> Dengyun Peng<sup>1</sup> Hanjing Li<sup>1</sup> Jiannan Guan<sup>1</sup>  
 Zheng Yan<sup>1</sup> Jiaqi Wang<sup>5</sup> Mengkang Hu<sup>6</sup> Yantao Du<sup>7</sup> Zhi Chen<sup>7</sup>  
 Xie Chen<sup>8</sup> Wanxiang Che<sup>1†</sup>

<sup>1</sup> Harbin Institute of Technology    <sup>2</sup> Central South University

<sup>3</sup> University of Illinois Urbana-Champaign    <sup>4</sup> Princeton University

<sup>5</sup> The Chinese University of Hong Kong    <sup>6</sup> The University of Hong Kong

<sup>7</sup> ByteDance Seed (China)    <sup>8</sup> Shanghai Jiao Tong University

{qgchen,jhliu,car}@ir.hit.edu.cn, qinlibo@hit.edu.cn

## Abstract

Understanding how information is dynamically accumulated and transformed in human reasoning has long challenged cognitive psychology, philosophy, and artificial intelligence. Existing accounts, from classical logic to probabilistic models, illuminate aspects of output or individual modelling, but do not offer a unified, quantitative description of general human reasoning dynamics. To solve this, we introduce Information Flow Tracking (IF-Track), that uses large language models (LLMs) as probabilistic encoder to quantify information entropy and gain at each reasoning step. Through fine-grained analyses across diverse tasks, our method is the *first successfully models the universal landscape of human reasoning behaviors* within a single metric space. We show that IF-Track captures essential reasoning features, identifies systematic error patterns, and characterizes individual differences. Applied to discussion of advanced psychological theory, we first reconcile single- versus dual-process theories in IF-Track and discover the alignment of artificial and human cognition and how LLMs reshaping human reasoning process. This approach establishes a quantitative bridge between theory and measurement, offering mechanistic insights into the architecture of reasoning.

**Key Words:** Human Reasoning Modelling, Information Theory, Cognitive Modeling, Large Language Models, Cognitive Psychology

## 1 Introduction

Human reasoning modelling has long been central to cognitive psychology, philosophy, and artificial intelligence, addressing how reasoning processes are structured [87, 91, 92, 48, 34, 99, 15]. Early approaches grounded in classical logic, such as propositional and deductive reasoning, modeled cognition through fixed rules and formal structures [87, 85]. With the rise of probabilistic paradigms, attention shifted to models incorporating heuristic and uncertainty-based mechanisms, exemplified by Bayesian reasoning frameworks [69, 11, 52, 86]. In contrast, the theory of mental models posits that reasoning involves constructing and manipulating possible situations rather than adhering to static rules [53]. Recent works refine this view, emphasizing reasoning as the evaluation of possibilities and necessities beyond formal logic [54]. More recently, meta-learning approaches highlight reasoning as an adaptive

---

\*Equal Contribution

†Corresponding Author

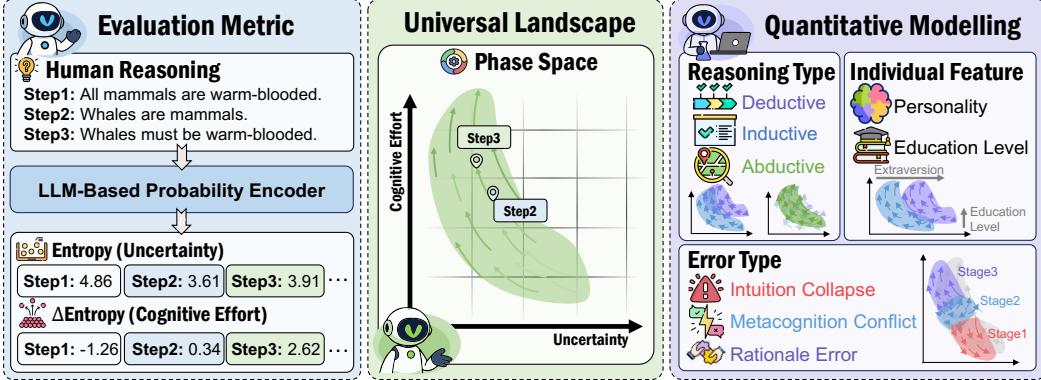


Figure 1: Theoretical Framework and Modelling Applications of IF-Track.

**a.** Computation of reasoning metrics. An LLM-based probability encoder estimates the conditional probability of each reasoning step, yielding two variables: uncertainty (information entropy,  $u_t$ ) and cognitive effort (information gain,  $e_t$ ). **b.** Theoretical foundation: Information phase space. Reasoning can be modeled as a trajectory within a two-dimensional information phase space ( $u_t, e_t$ ), formulated as a Hamiltonian system that models the reasoning landscape. This framework depicts the transition from high uncertainty and low effort (Step 2) to low uncertainty and high effort (Step 3). **c.** Applications under the framework. Built on this framework, IF-Track models reasoning across Reasoning Type (e.g. Deductive and Inductive Reasoning), Individual Feature (e.g., personality, professional background), and Error Type (e.g., intuition collapse, metacognitive conflict, rationale error).

process, where cognitive abilities evolve through interaction with the environment, providing a dynamic account of individual human cognition [6, 65, 77, 8].

From a neuroscience perspective, reasoning engages specific brain regions. Evidence indicates that human reasoning activates the prefrontal and parietal cortices, with theta oscillations in the prefrontal cortex linked to thought processes [2]. When reasoning output is correct, beta activity in EEG recordings increases significantly [81], providing biological support for reasoning models and underscoring the central role of neural dynamics in reasoning. Nevertheless, most current approaches focus on endpoint performance and isolated measures, lacking a unified quantitative framework to track general reasoning trajectories continuously. This limitation constrains mechanistic, process-level insight and obscures temporally resolved features, including the errors, types, and individual features of dynamic reasoning landscapes.

To address this gap, as illustrated in Fig. 1a, we introduce a framework that first treats large language models (LLMs) as probabilistic encoders to quantify the landscape of human reasoning by tracking, step-by-step along reasoning trajectories, information entropy (uncertainty), and the relevance development value (cognitive effort). Furthermore, as shown in Fig. 1b, to our knowledge, **Information Flow Tracking (IF-Track) framework delivers the first universal landscape of human reasoning** across diverse tasks through a unified information phase space. This formulation produces reproducible, cross-task signatures of cognitive landscape, enabling precise comparisons of reasoning strategies and uncovering previously inaccessible patterns of information flow. These insights offer substantial implications for understanding human cognition in societal contexts and advancing methodologies in cognitive and social sciences.

Further, as shown in Fig. 1c, we present a unified modelling framework that captures distinct reasoning patterns and models stepwise errors, showing how errors in intermediate states shape subsequent stages. Beyond feature modelling, it also models stable individual signatures, revealing how personality and educational background influence information processing and reasoning paths. These offer theoretical guidance for dissecting human reasoning behaviors and practical strategies for refining reasoning in large models. We apply this framework to key debates in cognitive psychology, including single- versus dual-process theories, which diverge locally yet converge globally. Concurrently, we show how LLMs reshape human reasoning, yielding insights on the evolution between human cognition and AI.

In summary, our contributions are as follows:

- **Providing Universal Landscape:** To the best of our knowledge, we are the first to quantitatively model universal human reasoning landscape, providing a new framework for quantitative analysis of reasoning behavior.
- **Effectively Modelling Reasoning Features:** We effectively capture and model key features of human reasoning processes, representing inductive and deductive reasoning in two distinct modes and integrating abductive reasoning through their combination.
- **Successfully Analyzing Individual Differences:** We quantify behavioral variation across individuals differing in personality and professional background, providing fresh insights into how such factors shape information processing and path selection.
- **Quantitative Application of Psychological Theory:** We apply this framework to discussions of psychological theories, such as single- versus dual-process models of reasoning, which differ locally but align globally. In parallel, we contrast how LLMs reshape human reasoning, yielding insights for aligning human cognition with AI.

## 2 Theory Model

To model the landscape of human reasoning, this section introduces the theoretical foundations of the **Information Flow Tracking** (IF-Track) framework, which model information changes governed by Hamiltonian dynamics. This approach demonstrates stable information flow during reasoning, analogous to physical systems. See Sec. 5.1 for a detailed description.

**Hamiltonian Dynamics** Generally, the evolution of a physical system can be elegantly formulated within **Hamiltonian dynamics** [56, 58, 3]. At any given time step  $t$ , the system’s state is characterized by a pair of conjugate variables, the *generalized coordinate*  $q_t$  and the *generalized momentum*  $p_t$ . Their temporal evolution follows Hamilton’s canonical equations:

$$\dot{q}_t = \frac{\partial H}{\partial p_t}, \quad \dot{p}_t = -\frac{\partial H}{\partial q_t}, \quad (1)$$

where  $H(q_t, p_t)$  denotes the Hamiltonian function, typically representing the total energy of the system. This formulation describes a *conservative flow* in which total energy is preserved and the trajectory evolves deterministically in the phase space. Beyond physical systems, in information phase space, uncertainty and cognitive effort can be modeled as conjugate variables reflecting informational energy and mental motion [50, 59, 32].

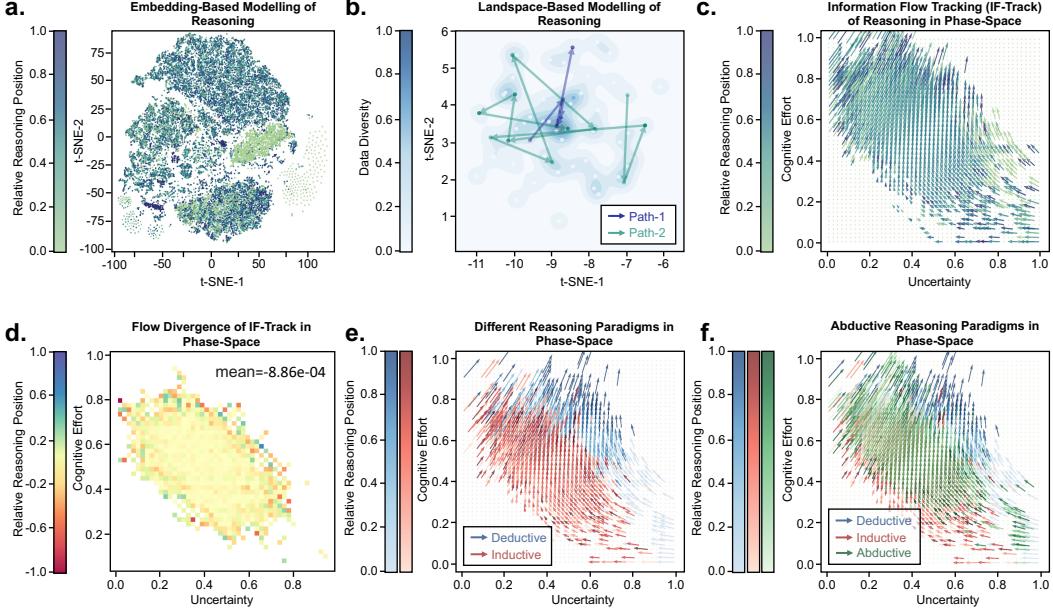
**Information Flow Tracking** Further, we define **Information Flow Tracking** (IF-Track) within an **information phase space** to describe reasoning as a continuous cognitive flow based on Hamiltonian dynamics. As shown in Fig. 1a, given each reasoning step  $t$ , the cognitive state is represented by a pair  $(u_t, e_t)$ , where  $u_t$  denotes the *uncertainty* (quantified by information entropy) and  $e_t$  denotes the *cognitive effort* (measured by the change in entropy, i.e., information gain between steps). The reasoning dynamics thus form a trajectory  $\mathbf{X}_t = (u_t, e_t)$ , which evolves according to an underlying information flow field  $\dot{\mathbf{X}}_t = \mathbf{V}(u_t, e_t)$ .

As shown Fig. 1b, each reasoning process can be viewed as a transition between two cognitive states (*Step1* → *Step2*) in this two-dimensional space: it typically begins in a high-uncertainty, low-effort region ( $u_1$  high,  $e_1$  low), reflecting intuitive exploration; and moves toward a low-uncertainty, high-effort region ( $u_2 < u_1$ ,  $e_2 > e_1$ ), representing deliberate analysis. This transition mirrors a *conservative evolution* in which total “informational energy” remains constant but redistributes between uncertainty and effort, an information-theoretic analogue to energy-momentum exchange in Hamiltonian dynamics.

**Information Flow Tracking meets Hamiltonian Dynamics.** According to **Liouville’s theorem**, Hamiltonian flow in phase space is divergence-free:

$$\nabla \cdot \vec{V} = \frac{\partial \dot{q}_t}{\partial q_t} + \frac{\partial \dot{p}_t}{\partial p_t} = 0, \quad (2)$$

implying that the total phase-space volume remains conserved during evolution [3]. We extend this conservation principle to human reasoning in IF-Track: Within the *information*



**Figure 2: Comparison of static representations of reasoning trajectories and relevant reasoning paradigms modelling.**

**a.** *t-SNE projection of step embeddings, showing chaotic representations across different human reasoning processes.* **b.** *Visualization of the landscape of thought, showing clustered but temporally unordered patterns across different human reasoning processes.* **c.** *Reasoning trajectories in the entropy–gain phase space reveal a globally consistent flow, where arrows represent the direction of information flow, illustrating the dynamic evolution of reasoning states from high uncertainty toward stable cognitive states.* **d.** *Empirical validation of the information phase space. The pseudocolor map shows the local divergence  $\nabla \cdot \vec{V}$  of the reasoning flow in the  $(u, e)$  space. Most regions exhibit near-zero divergence (uniform color), indicating approximate volume preservation and supporting a quasi-Hamiltonian structure of human reasoning dynamics.* **e.** *Different reasoning paradigms (deductive vs. inductive reasoning) in phase space.* **f.** *Abductive reasoning paradigms in phase space, lying between deductive and inductive and showing a hybrid pattern.*

phase space defined by uncertainty  $u_t$  and cognitive effort  $e_t$ , the reasoning information flow field  $\vec{V}(u_t, e_t) = (\dot{u}_t, \dot{e}_t)$  satisfies:

$$\nabla \cdot \vec{V}(u_t, e_t) = 0, \quad (3)$$

indicating that human reasoning maintains a **conserved structure in its information dynamics** (viz, **information flow field is phase space**). Empirically, it manifests in the smooth, divergence-free trajectories, where reasoning evolves continuously from intuitive to analytical states without loss of information. See more theoretical analyses in the Sec. 5.1.

### 3 Results

#### 3.1 Universal Human Reasoning Landscape Modelling

In this section, we validate the core theoretical claims of the IF-Track framework introduced in Sec. 2 by empirically testing whether it can successfully quantify and track human reasoning dynamics as an approximately incompressible information flow in phase space.

**Current static modelling methods can not unify modelling human reasoning landscape.** Most existing reasoning modelling methods emphasize static semantic distributions or output-only results, offering static snapshots rather than a dynamic process of reasoning [30, 95, 93]. As depicted in Fig. 2a, embedding the reasoning steps enables static visualizations of reasoning representations. Projecting these embeddings with t-SNE [63] produces clustered patterns but removes temporal order, obscuring how the thought process unfolds [9, 74]. The more advanced “landscape of thought” approach [98] renders these static representations as

smooth surfaces for multiple-choice tasks; however, for general reasoning, as shown in Fig. 2*b*, trajectories (green and purple) vary widely across problems, which undermines both the consistency and the interpretability of the modelling. Hence, neither their sequential order nor shared structure can be captured by those two static modelling methods, because the visualizations remain irregular and overlapping.

**Human reasoning process can be successfully quantified and tracked by IF-Track.** By mapping reasoning steps to the normalized phase space, as shown in Fig. 2*c*, IF-Track establishes an "information phase space" where the indicated arrow represents a consistent flow direction. This approach maintains coherent flow that enables both progression and interpretability. In contrast, non-reasoning scenarios (Fig. 7 in Methods) exhibit disordered dynamics. Thus, IF-Track quantifies reasoning trajectories as structured flow fields with consistent information paths. Notably, these flows show distinct dynamics: uncertainty decreases as intermediate conclusions accumulate, rebounding slightly at the final integrative step due to synthesis-induced doubt or further exploration; meanwhile, cognitive effort rises steadily. Overall, IF-Track integrates reasoning into a unified dynamic framework that captures the temporal directionality and structural essence of human thought.

**Human reasoning as an approximately incompressible information flow satisfying Liouville's equation in phase space.** To test whether the inferred cognitive dynamics are quasi-Hamiltonian and consistent with Liouville's equation, we computed the local divergence  $\nabla \cdot \vec{V}$  along reasoning trajectories and visualized it as a pseudocolor map in Fig. 2*d*. Extended regions of nearly uniform color indicate near-zero divergence (as yellow color in the Figure; mean < 1e-3), consistent with an approximately volume-preserving flow in phase space. Small deviations appear primarily at the boundaries, suggesting weak dissipative effects attributable to noise or boundary interactions. These observations support the theoretical soundness of our phase-space modelling of reasoning trajectories. Overall, the evidence is consistent with a quasi-Hamiltonian description in which uncertainty and cognitive effort act as effective conjugate variables that trade off while approximately conserving phase-space volume.

### 3.2 Reasoning Classical Attribution Modelling

To evaluate whether IF-Track can model classical attribution, we analyze both distinct reasoning types and common reasoning errors. Our framework shows capability in both respects: (1) it distinguishes classical reasoning types via trajectory patterns (Sec. 3.2.1); (2) it identifies and classifies reasoning errors as deviations from typical trajectories (Sec. 3.2.2).

#### 3.2.1 IF-Track Distinguishes Classical Reasoning Types via Trajectory Patterns.

Human reasoning is traditionally classified into three fundamental types in cognitive psychology [72, 43, 61, 80, 23]: (1) **Deductive reasoning** derives conclusions that necessarily follow from explicit premises, guaranteeing truth when the premises are true. (2) **Inductive reasoning** generalizes from specific observations to broader principles, yielding probabilistic rather than certain conclusions. (3) **Abductive reasoning** infers the most plausible explanations for incomplete or surprising evidence, supporting hypothesis generation and diagnostic inference. While these categories have long guided qualitative studies, whether IF-Track can model dynamic distinctions quantitatively is still an important question.

**Deductive and Inductive reasoning exhibit similar global patterns but distinct local dynamics.** As shown in Fig. 2*e*, both deductive and inductive reasoning follow a similar global pattern: uncertainty drops sharply at the outset, stabilizes midway, and slightly rebounds near the end, while cognitive effort rises steadily throughout. Yet their local dynamics diverge. Deductive reasoning starts with higher cognitive effort and rapid uncertainty reduction, consistent with its rule-based, top-down character. Inductive reasoning, by contrast, begins with lower effort and slower uncertainty reduction, reflecting exploratory pattern discovery. This difference supports the cognitive view that deduction and induction share a common structure but differ in their early-stage processing dynamics.

**Abductive Reasoning works with a hybrid dynamic pattern.** As illustrated in Fig. 2*f*, abductive reasoning occupies an intermediate position between deduction and induction on both uncertainty and effort. Its global trend mirrors the other two: uncertainty declines

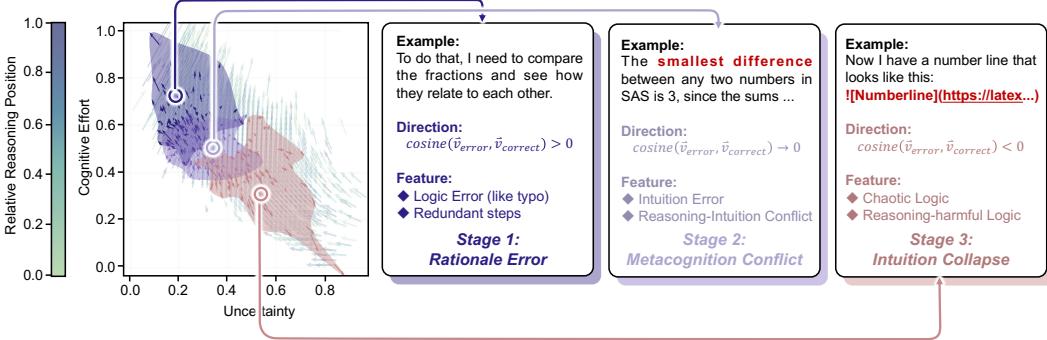


Figure 3: Three categories of reasoning errors identified by IF-Track, positioned in the uncertainty–effort phase space. ( $n_{\text{total}} = 9,991$ ,  $n_{\text{error}} = 374$ ). These errors were clustered into three stages by directions: *Stage 1: Intuition Collapse* (Ratio 87.3%), *Stage 2: Metacognition Conflict* (Ratio 73.7%), and *Stage 3: Rationale Error* (Ratio 90.4%)

and then slightly rebounds, while cognitive effort accumulates. Early steps show moderate uncertainty and low effort, consistent with tentative hypothesis formation. Subsequent steps alternate between surges in effort and shifts in uncertainty, reflecting iterative hypothesis testing and refinement. This hybrid trajectory presents abduction as a synthesis of exploratory inference (as in induction) and confirmatory reasoning (as in deduction), consistent with classical accounts of abductive cognition [71, 55, 51, 64].

### 3.2.2 IF-Track effectively identifies reasoning errors based on trajectory deviations.

Beyond classifying reasoning categories, a robust modelling approach must evaluate step-level correctness. To assess this capability, we analyze about 9,991 human-annotated reasoning steps that include 372 annotated erroneous steps across multiple categories. As shown in Fig. 3, these errors cluster into three stages consistent with Pennycook’s three-stage theory of reasoning errors [73]. Each stage maps to a distinct region defined by spatial position and directional dynamics of the reasoning trajectory, enabling IF-Track to identify error types from trajectory signatures. Specifically, these stages show the following features:

- **Stage 1: Intuition Collapse** is located in the lower-right corner of the phase space and marking the start of the reasoning flow with high uncertainty and low cognitive effort. Trajectories are impulsive and disorganized, often reversing direction ( $\cos(\vec{v}_{\text{error}}, \vec{v}_{\text{correct}}) < 0$ ), indicating motion opposite to the correct trajectory. These errors arise from faulty or unfounded intuition, where reasoning collapses before monitoring or deliberation.
- **Stage 2: Metacognition Conflict** is positioned in the central band of the phase space with moderate uncertainty and effort. It captures reasoning that appears coherent but rests on flawed assumptions. The direction cosine ( $\cos(\vec{v}_{\text{error}}, \vec{v}_{\text{correct}}) \approx 0$ ) indicates lateral divergence from the correct flow rather than reversal. Such errors reflect conflict-monitoring failures, where inconsistencies or contradictions go unnoticed during mid-stage reasoning.
- **Stage 3: Rationale Error** can be found in the upper-left area with low uncertainty but high cognitive effort. Reasoning remains aligned with the correct trajectory ( $\cos(\vec{v}_{\text{error}}, \vec{v}_{\text{correct}}) > 0$ ) yet suffers from inefficient or minor error processing, such as redundancy, over-explanation, or arithmetic slips, after the correct structure is established.

Together, these categories show that reasoning errors are systematically distributed across the uncertainty–effort phase space. This alignment provides empirical support that IF-Track not only detects reasoning failures but also recapitulates the cognitive progression described in human reasoning theories.

### 3.3 Effective Individual Characteristics Modelling

Beyond characterizing general reasoning patterns, our framework captures *individual variability* in reasoning behavior. We collected 6,452 reasoning trajectories from participants

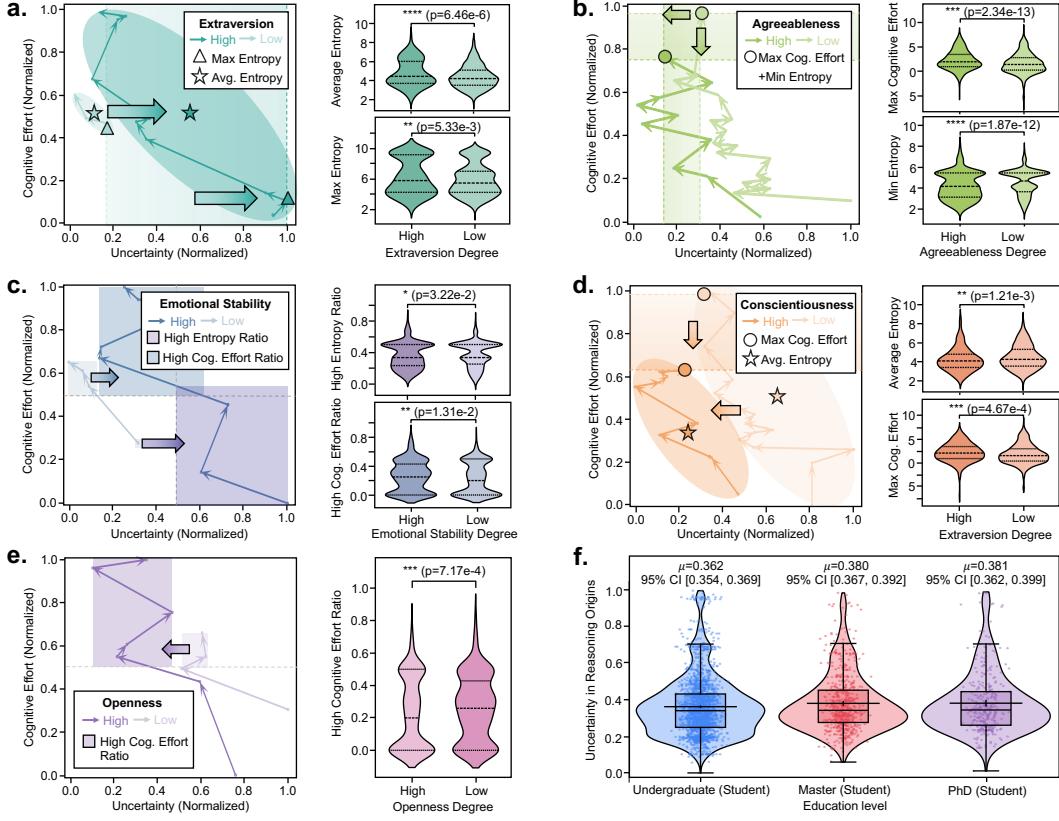


Figure 4: Personality-related modulation of reasoning trajectories and cognitive-informational dynamics.

**a. Extraversion.** Individuals with higher extraversion exhibit greater mean and maximal entropy, indicating higher tolerance for uncertainty and broader exploratory reasoning. **b. Conscientiousness.** The high-conscientiousness group shows lower average entropy but higher maximal effort, consistent with disciplined and goal-oriented reasoning. **c. Emotional Stability.** Individuals with higher emotional stability maintain a more balanced entropy–effort profile, reflected in higher ratios of high-entropy and high-effort states. **d. Openness.** Higher openness corresponds to a greater proportion of high-effort reasoning phases, suggesting deeper cognitive engagement and flexible information integration. **e. Agreeableness.** High-agreeableness participants show higher maximal cognitive effort and lower minimal entropy, suggesting more sustained and focused reasoning. **f. Education Level.** Across undergraduate, master, and PhD groups, higher educational attainment correlates with slightly higher uncertainty at reasoning origins, indicating broader hypothesis search spaces in early-stage reasoning. **Total**  $n = 3,215$ .

worldwide, spanning deductive, inductive, and abductive tasks. The dataset integrates demographic and psychological attributes, with emphasis on personality traits and education level, to test whether IF-Track quantitatively reveals individual characteristics.

### 3.3.1 Personality Traits Modelling

Personality traits modulate how individuals engage with **Uncertainty** and **Cognitive Effort** during reasoning. In this study, we adopt the Big Five Personality Traits as a comprehensive framework to characterize individual differences. This model captures five relatively independent dimensions, *Extraversion*, *Conscientiousness*, *Agreeableness*, *Emotional Stability*, and *Openness*, that jointly describe variations in affective tendencies, motivation, and cognitive processing.

**Extraversion: Preference for High-Uncertainty Exploration.** Arousal theory [28] posits that extraverts have lower baseline cortical arousal and therefore seek stimulation. As shown in Fig. 4a, individuals high in Extraversion exhibit higher average uncertainty ( $p = 6.46 \times 10^{-6}$ )

and maximal Uncertainty ( $p = 5.33 \times 10^{-3}$ ), consistent with a preference for ambiguous or unpredictable states and with evidence that extraversion and positive mood sustain persistence under ambiguity [47].

**Agreeableness: Seeking Efficient and Certain Trajectories.** In Fig. 4b, higher Agreeableness is associated with greater maximal Cognitive Effort ( $p = 2.34 \times 10^{-13}$ ) and lower maximal Uncertainty ( $p = 1.87 \times 10^{-12}$ ). Participants higher in Agreeableness tend to initiate reasoning in stable, low-Uncertainty states and then increase Cognitive Effort along structured trajectories, consistent with certainty-seeking, consensus-oriented, and conflict-averse processing characteristic of this trait [45].

**Emotional Stability: High Uncertainty Tolerance and Efficient Reasoning.** As shown in Fig. 4c, higher Emotional Stability is associated with higher proportions of high-uncertainty and high-cognitive-effort states ( $p < 0.05$ ). This pattern is consistent with evidence that individuals with high Emotional Stability tolerate uncertainty and ambiguity while maintaining coherent and efficient reasoning [31].

**Conscientiousness: Structured, Goal-Oriented Reasoning.** In Fig. 4d, participants high in conscientiousness show lower average uncertainty and higher peak cognitive effort ( $p < 0.01$ ). These individuals follow more structured reasoning paths, exhibiting reduced variability and uncertainty while concentrating peak cognitive effort at critical junctures [18].

**Openness: Greater Exploratory Engagement.** Fig. 4e indicates that higher Openness scores are associated with a greater proportion of high cognitive-effort states ( $p = 7.17 \times 10^{-4}$ ), while Uncertainty remains stable. This pattern suggests that individuals higher in Openness explore a broader range of reasoning paths without increasing Uncertainty [66, 22, 82].

### 3.3.2 Educational Level Modelling

**Educational attainment appears to shape not only knowledge but also the initial conditions of reasoning.** Fig. 4f compares the uncertainty of the first reasoning step across undergraduate, master's, and PhD participants. Higher education levels correspond to greater **initial uncertainty** ( $\mu_{\text{PhD}} = 0.381 > \mu_{\text{Undergrad}} = 0.362$ ), and the 95% confidence intervals overlap, indicating a consistent improvement. This pattern suggests that advanced academic training may encourage reasoning from broader hypothesis spaces, with reduced reliance on prior knowledge, greater self-directed exploration, and a higher tolerance for ambiguity at early stages [10, 84].

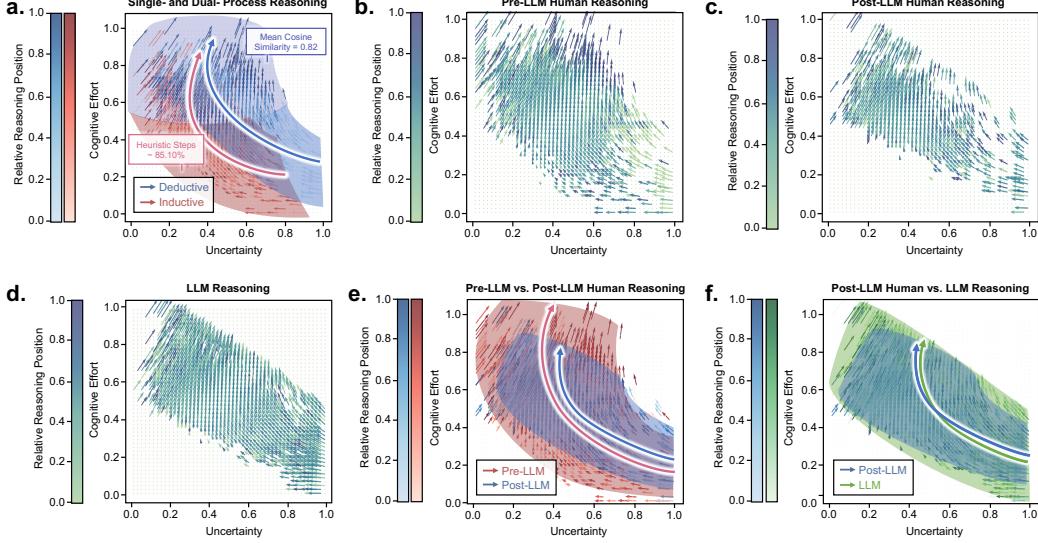
## 3.4 Application IF-Track for Advanced Psychological Theory Discussion

Based on previous analysis, we robustly conceptualise human reasoning as dynamic information flow in IF-Track within an information phase space. Building on this, we apply IF-Track to psychological theory, including dual-process accounts and human-LLM alignment.

### 3.4.1 Single- vs. Dual-Process Theories Debates

The long-standing debate over **single- vs. dual-process theories** of reasoning has focused on whether deductive and inductive reasoning stem from distinct systems or a single, continuous mechanism [24, 7]. Previous studies are inconclusive, some suggest separate neural activations, while others point to their integration and smooth transitions across reasoning stages [38, 19, 25]. Here, we argue that this apparent dichotomy can be reconciled through a unified account of reasoning dynamics offered by IF-Track. As shown in Fig. 5a, IF-Track positions intuitive and analytic modes within a single information-flow continuum.

**Locally, reasoning exhibits dual-process dynamics.** Inductive trajectories originate in high-Uncertainty, low-Effort regions (consistent with heuristic exploration) and evolve toward low-Uncertainty, high-Effort states (associated with analytic integration). Empirically, within the low-Effort regions identified by IF-Track, **85.10% of reasoning steps** can be manually classified as heuristic, indicating a dominant intuitive phase during early reasoning. This intra-episode shift from intuitive to deliberate processing quantitatively captures dual-process phenomena within individual reasoning paths [26, 27, 68].



**Figure 5: Application IF-Track for Advanced Psychological Theory Discussion.**

**a.** Comparison between single- and dual-process theories of reasoning ( $n = 1632$ ). Dual-process theories posit interacting intuitive and analytic systems, whereas single-process accounts treat reasoning as a graded, continuously integrated computation. **b.** Pre-LLM human reasoning flow ( $n = 1667$ ). Before the LLM era, human reasoning often began with low effort under high uncertainty and, via analytic integration, progressed toward higher effort with lower uncertainty. **c.** Model reasoning flow ( $n = 1537$ ). LLMs exhibit a similar trajectory: uncertainty declines as computational depth increases, mirroring human analytic patterns. **d.** Post-LLM human reasoning flow ( $n = 1549$ ). After interacting with LLMs, human reasoning shows a compressed trajectory, starting at higher effort and lower uncertainty, converging sooner, and often skipping further exploration. **e.** Comparison of pre- and post-LLM human reasoning. Pre-LLM reasoning featured extended exploration and late convergence, whereas post-LLM reasoning stabilizes earlier and is more efficient, signaling a shift from discovery-oriented to synthesis-oriented cognition. **f.** Comparison between post-LLM human and LLM reasoning. Both trajectories now begin at similar levels of uncertainty and effort, suggesting an emerging convergence in reasoning structure across humans and models.

**Globally, reasoning follows a single-process flow.** Aggregated across tasks and participants, reasoning trajectories exhibit a consistent, monotonic decrease in Uncertainty and a steady increase in Cognitive Effort. When comparing later-stage reasoning across deductive and inductive datasets, the **mean cosine similarity between their trajectory vectors reached 0.82**, suggesting strong alignment and structural consistency in the global reasoning flow. This large-scale regularity supports a single-process framework, indicating that heuristic and analytic modes are dynamically coupled components of a unified system [20, 68]. Thus, IF-Track shows that dual-process effects arise as local transitions within a global single-process architecture.

### 3.4.2 Human Reasoning Reshaping in the Era of LLMs

With the rapid development of LLMs, they are increasingly used as essential tools in daily life and work, supporting reasoning and decision-making. Specifically, we utilize GPT-4o [1] as a strong model provided stable reasoning. Concurrent studies indicate that frequent reliance on AI tools can reshape human behavior distributions [37, 76, 33, 79, 36, 83, 5, 16]. This raises a key question: **To what extent does reliance on GPT-4o for reasoning lead users to implicitly adopt the model’s reasoning patterns, thereby altering their subsequent reasoning in the model’s absence?** The resulting patterns are depicted in Fig. 5b–f.

**LLMs are reshaping human reasoning.** As shown in Fig. 5b–e, pre-LLM reasoning typically begins with low cognitive effort, reflecting tentative initial intuitions, and increases through exploration and iterative refinement, yielding a low-start, high-end trajectory. In contrast, with extensive reliance on LLMs, reasoning often starts at a higher level of cognitive effort but

tends to end lower, producing a high-start, lower-end trajectory. Intuitively, the former builds effort through open-ended search, whereas the latter proceeds along a more constrained pathway that dampens later-stage exploratory effort. Together, these patterns indicate a redistribution of cognitive effort across stages of the reasoning process in the LLM era.

**Post-LLM human reasoning flows closely align with those of LLMs.** As shown in Fig. 5f, trajectories produced by GPT-4o largely overlap with human reasoning flows after its release. This overlap suggests that frequent LLM use not only changes the context in which people reason but also subconsciously encourages users to mimic and internalize model-specific heuristics, promoting convergence between human and machine reasoning. Post-LLM human trajectories exhibit high initial Cognitive Effort followed by low terminal effort, narrowing the accessible region of the reasoning phase space. Individuals, aligned with LLMs, appear less inclined toward prolonged exploration, and the terminal segment of information flow loses the previously observed exploratory region characterized by reduced uncertainty and elevated Cognitive Effort.

## 4 Conclusion & Discussion

In summary, we present a unified, stepwise framework that quantitatively captures the dynamics of human reasoning by tracing information entropy and gain through inferential trajectories. Our approach reconciles classical and probabilistic theories, formalizes reasoning processes in measurable terms, and uncovers individual and group-level cognitive signatures. By applying these tools to discussions on single- versus dual-process theories and comparing human with large language model reasoning, we provide new views for aligning AI with human thought and quantify how LLMs reshape human reasoning.

Future work could extend this framework to real-time neural recordings and dynamic decision-making contexts to further elucidate the neurocognitive mechanisms underlying reasoning. Moreover, IF-Track may enable the application of this approach in adaptive cognitive training paradigms, allowing for the assessment and enhancement of reasoning skills in educational and clinical settings.

## 5 Method

This section presents the methodological framework and implementation of our study, which quantitatively models human and model reasoning trajectories within the information phase space defined by **uncertainty** and **cognitive effort**.

### 5.1 Detailed IF-Track Framework

In this section, we elaborate on the IF-Track framework introduced in Sec. 2, detailing how it quantitatively computes uncertainty and cognitive effort, and how it formalizes reasoning dynamics as a Hamiltonian system within the information phase space.

#### 5.1.1 Quantifying Uncertainty and Cognitive Effort

We next describe in detail how IF-Track quantitatively computes **uncertainty** and **cognitive effort** for each reasoning trajectory. These two quantities form the orthogonal dimensions of the information phase space, representing respectively the ambiguity of reasoning states and the cognitive adjustment required between consecutive steps.

**Uncertainty.** Given a reasoning step containing  $n$  tokens with probabilities  $\{p_i\}_{i=1}^n$ , the uncertainty of that step is defined as the average token-level Shannon entropy [29]:

$$u_t = -\frac{1}{n_t} \sum_{i=1}^{n_t} p_{t,i} \log p_{t,i}, \quad (4)$$

which reflects the model’s internal uncertainty when generating the  $t$ -th reasoning step.

**Cognitive Effort.** Cognitive effort is defined as the temporal derivative of uncertainty along the reasoning trajectory, representing the rate of entropy change between adjacent steps.

Expanding  $u_t$  and  $u_{t-1}$  from the above definition yields:

$$e_t = u_t - u_{t-1} = -\frac{1}{n_t} \sum_{i=1}^{n_t} p_{t,i} \log p_{t,i} + \frac{1}{n_{t-1}} \sum_{j=1}^{n_{t-1}} p_{t-1,j} \log p_{t-1,j}. \quad (5)$$

This formulation expresses cognitive effort as a difference in token-level entropy expectations between consecutive steps, that is, a reorganization of probability mass  $\{p_{t,i}\}$  along the reasoning sequence. It directly quantifies the magnitude of information restructuring required for cognitive progression, aligning with the concept of cognitive effort in cognitive science.

**Normalization.** To ensure comparability across reasoning trajectories, we employ two complementary normalization strategies.

1. *Global normalization:* Both uncertainty and cognitive effort are individually normalized across the entire dataset to the range  $[0, 1]$ , facilitating comparisons across different datasets or reasoning paradigms by aligning all measurements onto a shared scale.
2. *Local normalization:* For visualization and intra-sample comparison, each trajectory's step indices are linearly normalized to  $[0, 1]$ , preserving its internal temporal dynamics and enabling meaningful comparisons between reasoning steps within the same trajectory.

Since homeomorphic transformations do not alter the topological structure of trajectories, these normalizations unify the measurement scales without distorting the underlying flow geometry. Such rescaling helps us observe the intrinsic dynamical patterns of reasoning trajectories while minimizing the influence of scale differences across datasets or individuals.

### 5.1.2 Liouville Conservation in the Information Phase Space

**Notation & Assumptions:** Let the reasoning process evolve continuously over the normalized reasoning step  $\tau \in [0, 1]$ . At each time  $\tau$ , the reasoning state is denoted as:

$$\mathbf{X}_\tau = (u_\tau, e_\tau), \quad (6)$$

where  $u_\tau$  represents *uncertainty* (information entropy) and  $e_\tau$  represents *cognitive effort* (information gain between consecutive steps). The reasoning dynamics is modeled as a continuous flow in the 2D information flow fields ( $\Omega \subset \mathbb{R}^2$ ):

$$\dot{\mathbf{X}}_\tau = \mathbf{V}(u_\tau, e_\tau) = (V_1(u_\tau, e_\tau), V_2(u_\tau, e_\tau)), \quad (7)$$

where  $V_1$  and  $V_2$  denote the instantaneous value of uncertainty and cognitive effort for human reasoning, respectively. That is,  $V_1 = \dot{u}_\tau$  and  $V_2 = \dot{e}_\tau$ .

Let  $\rho(u_\tau, e_\tau, \tau)$  denote the probability density of reasoning states in the information flow fields. Assuming that  $\rho$  are change smoothly, it formally has the following assumption:

**Assumption:** Under the *quasi-stationary* condition ( $\partial_\tau \rho \approx 0$ ) and assuming that  $\rho$  varies slowly over  $\tau$ , The density can be treated as locally time-invariant. This implies that the evolution of reasoning preserves local information volume in expectation.

**Continuity equation and Liouville condition.** Conservation of probability mass in phase space yields the continuity equation:

$$\frac{\partial \rho}{\partial \tau} + \nabla \cdot (\rho \mathbf{V}) = 0, \quad \text{where } \nabla \cdot (\rho \mathbf{V}) = \frac{\partial(\rho V_1)}{\partial u_\tau} + \frac{\partial(\rho V_2)}{\partial e_\tau}. \quad (8)$$

Under the above assumption,  $\partial_\tau \rho \approx 0$  and  $\nabla \rho \approx 0$ , which simplifies Eq. (8) to

$$\nabla \cdot \mathbf{V} = \frac{\partial V_1}{\partial u_\tau} + \frac{\partial V_2}{\partial e_\tau} = 0, \quad (9)$$

known as the **Liouville condition** for incompressible information flow.

**Measure preservation and Liouville's theorem.** Let  $\Phi_\tau : \Omega \rightarrow \Omega$  be the flow generated by a vector field  $\mathbf{V}$ , following Eq. (7). This function maps an initial point  $X_0$  to its new

position  $\Phi_\tau(X_0) = (f_1, f_2, \dots, f_n)$  after evolving for a time  $\tau$ . The derivative of this map is the Jacobian matrix,  $J_\tau$ :

$$J_\tau = D\Phi_\tau = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{pmatrix}, \quad (10)$$

which describes how the flow  $\Phi_\tau$  locally applies a linear transformation to the space, such as stretching, compression, or rotation. Its determinant,  $\det J_\tau$ , measures the factor by which the volume of an infinitesimal element changes after evolving for time  $\tau$ . The Liouville identity gives the time evolution of this volume change:

$$\frac{d}{d\tau} \log \det J_\tau = (\nabla \cdot \mathbf{V})(\Phi_\tau(\mathbf{X}_0)) \Rightarrow \det J_\tau = \exp\left(\int_0^\tau \nabla \cdot \mathbf{V}(\Phi_s(\mathbf{X}_0)) ds\right). \quad (11)$$

Hence,  $\det J_\tau = 1$  if and only if  $\nabla \cdot \mathbf{V} = 0$ , showing that Eq. (9) is equivalent to phase-space measure preservation, indicating that the information flow is volume-preserving.

### 5.1.3 Discretized Information Phase Space

**Hamiltonian representation of divergence-free flows.** In any simply connected two-dimensional domain, a continuously differentiable divergence-free vector field admits a scalar potential  $H(u_\tau, e_\tau)$  such that:

$$\mathbf{V} = \nabla^\perp H = \left( \frac{\partial H}{\partial e_\tau}, -\frac{\partial H}{\partial u_\tau} \right), \quad (12)$$

leading to the canonical Hamiltonian system:

$$\dot{u}_\tau = \frac{\partial H}{\partial e_\tau}, \quad \dot{e}_\tau = -\frac{\partial H}{\partial u_\tau}, \quad (13)$$

where  $H(u_\tau, e_\tau)$  remains conserved along trajectories, that meets:

$$\frac{dH}{d\tau} = \nabla H \cdot \dot{\mathbf{X}}_\tau = 0. \quad (14)$$

Thus, reasoning dynamics behaves as a Hamiltonian flow in  $(u_\tau, e_\tau)$  information flow fields.

**Simplified Hamiltonian Function under the information-theoretic constraint.** Given the empirical constraint that cognitive effort reflects the rate of change of uncertainty ( $e_\tau = \dot{u}_\tau$ ), we have  $V_1 = e_\tau$ . Substituting into Eq. (9) gives  $\partial_{e_\tau} V_2 = 0 \Rightarrow V_2 = -U'(u_\tau)$ . Integrating  $\partial_{e_\tau} H = e_\tau$  yields the separable Hamiltonian:

$$H(u_\tau, e_\tau) = \frac{1}{2}e_\tau^2 + U(u_\tau) + C, \quad \dot{u}_\tau = e_\tau, \quad \dot{e}_\tau = -U'(u_\tau), \quad (15)$$

for which  $\frac{dH}{d\tau} = e_\tau \dot{e}_\tau + U'(u_\tau) \dot{u}_\tau = 0$ . Hence,  $u_\tau$  and  $e_\tau$  constitute a conjugate pair of generalized coordinate and momentum in the information phase space.

**Finite-volume form ensures divergence discretization.** For a bounded domain  $\Omega = [0, 1] \times [0, 1]$ , let  $C_{ij}$  be a rectangular control cell centered at  $(u_i, e_j)$  with sizes  $(\Delta u, \Delta e)$ . By the divergence theorem,

$$(\nabla \cdot \mathbf{V})_{ij} = \frac{1}{|C_{ij}|} \oint_{\partial C_{ij}} \mathbf{V} \cdot \hat{\mathbf{n}} ds \approx \frac{1}{\Delta u} [\bar{u}_{i+\frac{1}{2},j} - \bar{u}_{i-\frac{1}{2},j}] + \frac{1}{\Delta e} [\bar{e}_{i,j+\frac{1}{2}} - \bar{e}_{i,j-\frac{1}{2}}], \quad (16)$$

where the edge-averaged fluxes are:

$$\bar{u}_{i\pm\frac{1}{2},j} = \frac{1}{\Delta e} \int_{e_j - \frac{\Delta e}{2}}^{e_j + \frac{\Delta e}{2}} \dot{u}(u_{i\pm\frac{1}{2}}, e) de, \quad \bar{e}_{i,j\pm\frac{1}{2}} = \frac{1}{\Delta u} \int_{u_i - \frac{\Delta u}{2}}^{u_i + \frac{\Delta u}{2}} \dot{e}(u, e_{j\pm\frac{1}{2}}) du. \quad (17)$$

Approximating edge integrals by the midpoint rule yields (Fig. 2c uses this discretization):

$$(\nabla \cdot \mathbf{V})_{ij} = \frac{1}{\Delta u} [\dot{u}_{i+\frac{1}{2},j} - \dot{u}_{i-\frac{1}{2},j}] + \frac{1}{\Delta e} [\dot{e}_{i,j+\frac{1}{2}} - \dot{e}_{i,j-\frac{1}{2}}] + O(\Delta u^2 + \Delta e^2), \quad (18)$$

which provides the standard second-order finite-volume discretization of the Liouville condition in information phase space.

### Summary

- Probability conservation  $\Rightarrow$  continuity equation (Eq. (8));
- Measure preservation  $\Leftrightarrow$  divergence-free condition (Eq. (9));
- In 2D, divergence-free  $\Leftrightarrow$  Hamiltonian structure (Eq. (12));
- Under  $e_\tau = \dot{u}_\tau$ , Hamiltonian function  $H(u_\tau, e_\tau) = \frac{1}{2}e_\tau^2 + U(u_\tau)$  (Eq. (15));
- Finite-volume form (Eqs. (16)–(18)) ensures divergence discretization.

## 5.2 Experimental Setting

### 5.2.1 Experimental Setting under IF-Track.

All experiments conducted within our framework follow a unified modelling and data processing configuration. Each reasoning trajectory for IF-Track is encoded using the **Llama3-8B-Instruct** model, which transforms both the input problem and its step-by-step reasoning process into high-dimensional semantic representations for subsequent computation of information uncertainty and cognitive effort. Consistent normalization and feature extraction procedures are applied across all reasoning types to ensure comparability.

Unless otherwise specified, all experiments are performed under the same hardware environment and random seed settings to guarantee stability and reproducibility.

### 5.2.2 Embedding-based Visualization

To examine the geometric structure of reasoning in the information phase space, we compute stepwise embeddings for each trajectory using the “[CLS]” representation in BERT [21] model. Each step is encoded as a semantic vector that approximates its latent reasoning state. The resulting collection of embeddings defines a high-dimensional manifold that captures reasoning dynamics.

To visualize the manifold, we apply t-SNE [63] to project the embeddings into two dimensions (`random_seed=42`). The resulting map preserves local continuity across successive reasoning steps and reveals global clusters of reasoning patterns. A color gradient encodes each step’s position within its sequence, providing an interpretable view of how model uncertainty and a proxy for cognitive effort vary along the reasoning trajectory.

### 5.2.3 Landscape of Thought for Open-Ended Reasoning

To extend the original *Landscape of Thought* [98] framework, originally designed only for multiple-choice reasoning, to support open-ended reasoning tasks, we adapted the method through a unified sampling and transformation procedure. For each open-ended question, we sampled multiple human- or model-generated responses and constructed a **pseudo multiple-choice set**, where multiple sampled answers as choice set. This design enables consistent embedding and visualization of diverse reasoning trajectories within the same representational space. For re-implementation, specifically, each reasoning step within these sampled responses was embedded into a semantic vector using BERT. The resulting high-dimensional features were normalized and projected into a two-dimensional manifold using t-SNE to preserve both local continuity and global relational structure across steps.

We then applied a kernel density estimation to capture the overall distribution of reasoning trajectories, generating a continuous “cognitive landscape” that reflects areas of high reasoning convergence and exploratory dispersion. Representative trajectories were visualized by tracing their progression across the landscape, where transparency and color gradients encode step progression from early heuristic exploration to later analytic consolidation.

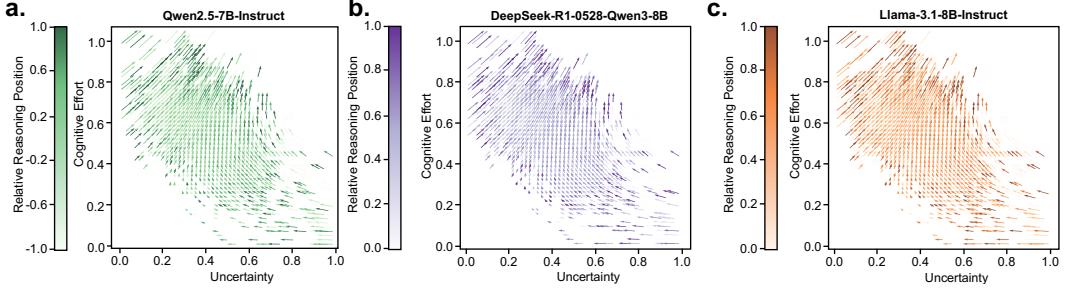


Figure 6: The generalization analysis of IF-Track on Qwen2.5-7B-Instruct (a), DeepSeek-R1-0528-Qwen3-8B (b), Llama-3.1-8B-Instruct (c).

#### 5.2.4 IF-Track on Non-Reasoning Scenarios

To validate the specificity of the IF-Track framework to reasoning processes, control experiments were conducted on non-reasoning tasks such as conversational dialogue tasks, text summarization, and machine translation. For these comparisons, the *Persona-Chat* [49] dataset was adopted to represent conversational and narrative text generation without explicit reasoning chains, allowing us to evaluate whether the proposed entropy-effort dynamics emerge only in genuine reasoning processes rather than general language modeling behaviors.

Using the same model and data processing pipeline as in the reasoning experiments, we computed uncertainty and cognitive effort for these tasks. We then analyzed the resulting trajectories in the information phase space to assess whether they exhibit the same Hamiltonian structure and conservation properties as observed in reasoning tasks. This comparison helps to elucidate whether the uncertainty-effort dynamics are unique to reasoning or represent a more general cognitive phenomenon.

#### 5.2.5 Generalization Analysis on Different LLMs

To assess the generalizability of the IF-Track framework across different LLMs, we replicated our experiments using multiple LLM architectures, including Qwen2.5-7B-Instruct [89], DeepSeek-R1-0528-Qwen3-8B [42], and Llama-3.1-8B-Instruct [41]. For each model, we followed the same data processing and analysis procedures as outlined in previous sections. We computed uncertainty and cognitive effort for reasoning trajectories generated by each model and examined their dynamics in the information phase space.

By comparing the results across different LLMs, we aimed to determine whether the flow structure and properties identified by IF-Track are consistent features of reasoning processes across diverse model architectures. As shown in Figure 6, the results demonstrate that all tested LLMs exhibit the absolutely same information change dynamics in their reasoning trajectories, supporting the significant robustness and universality of the IF-Track framework.

### 5.3 Data Collection

#### 5.3.1 The Collection of Comprehensive Human Reasoning Data

To comprehensively validate the generalizability of IF-Track, we construct an integrated reasoning dataset covering diverse domains and reasoning types. As illustrated in Table 1,

Dataset	Domain	Reasoning Type	Data Size
AIME2024 [70]	Mathematics	Deductive	30
GSM8K [17]	Mathematics	Deductive	8K
BigGSM [13, 14]	Mathematics	Deductive	610
MATH [46]	Mathematics	Deductive	12K
NuminaMathCoT [4]	Mathematics	Deductive	16K
OlympiadBench [44]	Mathematics / Science	Deductive & Inductive	9K
SciFact [90]	Science	Abductive	1.4K
PHYBench [75]	Physics	Inductive & Abductive	1K
WorldTree V2 [94]	Science	Deductive & Inductive	4.4K
CommonSenseQA [88]	Commonsense	Inductive & Abductive	9K
OpenBookQA [67]	Commonsense / Science	Deductive & Inductive	5K
AUQA [35]	Multimodal / Art	Inductive & Abductive	3K
LogiQA [62]	Logic	Deductive	1.8K
CRT-QA [96]	Critical Reasoning	Deductive & Inductive	728
LSAT (AGI-Eval [97])	Logic / Exam	Deductive	2K
Gaokao (AGI-Eval [97])	Examination	Deductive & Inductive	4K
JEC-QA (AGI-Eval [97])	Law / Examination	Deductive & Inductive	2K
EKAR [12]	Analogy	Inductive	1.1K
GPQA [78]	Graduate-level / Knowledge	Deductive & Abductive	209
PRM800K [60]	Process Supervision (multi-domain)	Deductive & Abductive	10K

Table 1: Overview of datasets used for comprehensive reasoning evaluation. Each dataset is categorized by its domain, reasoning type, and approximate data size (total  $\sim 112$ K samples).

this dataset includes more than 100,000 reasoning samples collected from a wide range of existing datasets, spanning mathematics, science, commonsense, logic, and examination-style reasoning. Each dataset contributes a distinct perspective on reasoning dynamics, enabling us to assess whether the uncertainty–effort framework holds consistently across different cognitive tasks.

These datasets collectively span mathematics, science, commonsense, logic, and human-level reasoning tasks, providing a comprehensive foundation for analyzing reasoning dynamics across domains. All datasets are unified into a consistent JSONL format, where each entry contains a question, a multi-step reasoning process, and a final answer. This unified structure allows us to extract token-level entropy for each reasoning step and compute the corresponding cognitive effort as defined in Section 5.1.1.

### 5.3.2 The Collection of Human Reasoning Data with Individual Features

To better understanding of modelling capabilities on individual features , we present the design and implementation of a large-scale study of human reasoning that captures detailed reasoning trajectories by free-text input alongside individual cognitive characteristics. Specifically, the study comprises two components: participant recruitment and questionnaire design.

**Participant Design.** We collected 6,452 reasoning trajectories by entrusting commercial companies from participants across 15 countries, with the geographical distribution shown in Figure 8a. The participant pool covered a wide range of educational backgrounds, from undergraduate to doctoral level. Nationality–education and nationality–gender distributions are visualized using chord diagrams in Figure 8 b,c, demonstrating the diversity of population.

**Questionnaire Design.** The major problems in the questionnaire are constructed based on the AGI-Eval [97] benchmark, covering multiple domains including mathematics, medicine, computer science, humanities, and history. To ensure balanced coverage of reasoning paradigms, we included three reasoning types: deduction, induction, and abduction. For each type, six representative problems were randomly sampled from 6K independently for each participant, and their order was randomized to control for sequence effects. In addition to task performance, we also collected personality information based on the Big Five dimensions, using the Ten-Item Personality Inventory (TIPI) [40], a concise 10-item measure that captures the Big Five traits (Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Openness). This enables us to analyze how psychological traits relate to reasoning dynamics.

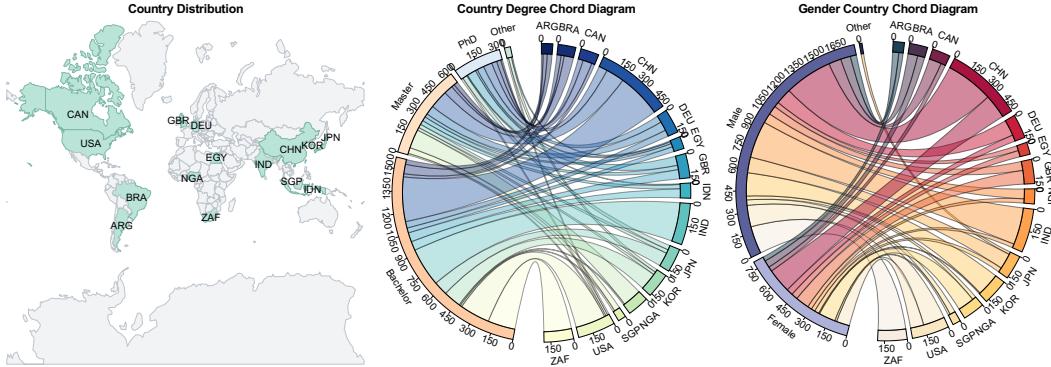


Figure 8: Geographical and demographic distribution of study participants, illustrating the distribution between nationality and education, and between nationality and gender.

### 5.3.3 The Collection of Pre-LLM and Post-LLM Human Reasoning Data

To examine the impact of LLMs on human reasoning, we implemented a two-phase data collection protocol: pre-LLM and post-LLM reasoning tasks.

**Pre-LLM Human Reasoning Data** In the pre-LLM phase, participants tackled a series of reasoning problems across diverse domains, including mathematics, science, and commonsense reasoning (drawn from the AGI-Eval benchmark). They solved these problems independently and recorded their step-by-step reasoning. To select problems predate the release of GPT-4o, we choose AGI-Eval to avoid data leakage. Moreover, we selected participants with no prior LLM experience.

**LLM Reasoning Data** In the LLM phase, we prompted the LLMs used in our study (e.g., GPT-4o [1]) with the same AGI-Eval questions. Following Kojima et al. [57], each prompt explicitly requested step-by-step reasoning (e.g., “Let’s think step-by-step!”). To capture both typical and diverse reasoning behaviors, we used standard decoding settings ( $\text{top-p}=0.95$ ,  $\text{temperature}=0.6$ ). Following Golovneva et al. [39], all model outputs were automatically segmented into reasoning steps and aligned with human step boundaries when available.

**Post-LLM Human Reasoning Data** In the post-LLM phase, pre-LLM phase participants first received regular exposure to GPT-4o (daily usage) via guided practice sessions. We then recruited the same cohort or a demographically matched group to revisit a subset of the original problems with similar difficulty and categories, avoiding exact duplicates to prevent knowledge leakage. Participants are also banned from the use of LLMs while documenting their step-by-step processes to show the change of the human reasoning process.

This two-phase design, applied to the same problem subset and comparable participant demographics, enabled direct comparison of human reasoning trajectories before and after LLM exposure. It thus provides insights into how LLMs influence human cognitive processes and reasoning patterns.

## 6 Ethical Considerations

All procedures involving human participants were reviewed, and informed consent was obtained before data collection. All data were anonymized to ensure participant privacy. Participants received fair compensation in accordance with institutional guidelines. A professional labeling company annotated the reasoning data at a rate of \$2.5 per participant, and all labelers held at least a college degree.

## 7 Code Availability

The code and relevant data are available at <https://github.com/LightChen233/Human-Reasoning-Modeling>.

## References

- [1] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.
- [2] Hafeez Ullah Amin, Aamir Saeed Malik, Muhammad Hussain, Nidal Kamel, and Weng-Tink Chooi. Brain behavior during reasoning and problem solving task: an eeg study. In *2014 5th International Conference on Intelligent and Advanced Systems (ICIAS)*, pages 1–4. IEEE, 2014.
- [3] Vladimir Igorevich Arnol'd. *Mathematical methods of classical mechanics*, volume 60. Springer Science & Business Media, 2013.
- [4] Edward Beeching, Shengyi Costa Huang, Albert Jiang, Jia Li, Benjamin Lipkin, Zihan Qina, Kashif Rasul, Ziju Shen, Roman Soletskyi, and Lewis Tunstall. Numinamath 7b cot. <https://huggingface.co/AI-MO/NuminaMath-7B-CoT>, 2024.
- [5] Emily M Bender and Alexander Koller. Climbing towards nlu: On meaning, form, and understanding in the age of data. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, pages 5185–5198, 2020.
- [6] Marcel Binz, Ishita Dasgupta, Akshay K Jagadish, Matthew Botvinick, Jane X Wang, and Eric Schulz. Meta-learned models of cognition. *Behavioral and Brain Sciences*, 47:e147, 2024.
- [7] Linda AW Brakel and Howard Shevrin. *Comment: Individual differences in reasoning: Implications for the rationality debate?* Cambridge University Press, 2003.
- [8] Rasmus Bruckner, Hauke R Heekeren, and Matthew R Nassar. Understanding learning through uncertainty and bias. *Communications Psychology*, 3(1):24, 2025.
- [9] Pedro C. Vieira, João P Montrezol, João T. Vieira, and João Gama. S+ t-sne-bringing dimensionality reduction to data streams, 2024.
- [10] Raymond Bernard Cattell. *Intelligence: Its structure, growth and action*, volume 35. Elsevier, 1987.
- [11] Nick Chater, Mike Oaksford, Ulrike Hahn, and Evan Heit. Bayesian models of cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(6):811–823, 2010.
- [12] Jiangjie Chen, Rui Xu, Ziquan Fu, Wei Shi, Zhongqiao Li, Xinbo Zhang, Changzhi Sun, Lei Li, Yanghua Xiao, and Hao Zhou. E-kar: A benchmark for rationalizing natural language analogical reasoning. 2022.
- [13] Qiguang Chen, Libo Qin, Jiaqi Wang, Jingxuan Zhou, and Wanxiang Che. Unlocking the capabilities of thought: A reasoning boundary framework to quantify and optimize chain-of-thought. *Advances in Neural Information Processing Systems*, 37:54872–54904, 2024.
- [14] Qiguang Chen, Libo Qin, Jinhao Liu, Yue Liao, Jiaqi Wang, Jingxuan Zhou, and Wanxiang Che. Rbf++: Quantifying and optimizing reasoning boundaries across measurable and unmeasurable capabilities for chain-of-thought reasoning, 2025.
- [15] Qiguang Chen, Libo Qin, Jinhao Liu, Dengyun Peng, Jiannan Guan, Peng Wang, Mengkang Hu, Yuhang Zhou, Te Gao, and Wanxiang Che. Towards reasoning era: A survey of long chain-of-thought for reasoning large language models. *arXiv preprint arXiv:2503.09567*, 2025.
- [16] Xusen Cheng and Lulu Zhang. Inspiration booster or creative fixation? the dual mechanisms of llms in shaping individual creativity in tasks of different complexity. *Humanities and Social Sciences Communications*, 12(1):1–10, 2025.
- [17] Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, et al. Training verifiers to solve math word problems. 2021.
- [18] Paul T Costa and Robert R McCrae. *The revised neo personality inventory (neo-pi-r)*, volume 2. 2008.

- [19] Adam L Darlow and Steven A Sloman. Two systems of reasoning: Architecture and relation to emotion. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(3):382–392, 2010.
- [20] Wim De Neys. On dual-and single-process models of thinking. *Perspectives on psychological science*, 16(6):1412–1427, 2021.
- [21] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019. URL <https://arxiv.org/abs/1810.04805>.
- [22] Colin G. DeYoung, Jordan B. Peterson, and Daniel M. Higgins. Exploring the hierarchy of personality: Integrating the big five and trait aspects of intelligence. *Personality and Individual Differences*, 33(4):533–552, 2002. doi: 10.1016/S0191-8869(01)00171-1.
- [23] Igor Douven. Abduction. 2011.
- [24] Jonathan St BT Evans. Dual-processing accounts of reasoning, judgment, and social cognition. *Annu. Rev. Psychol.*, 59(1):255–278, 2008.
- [25] Jonathan St BT Evans. *Thinking twice: Two minds in one brain*. Oxford University Press, 2010.
- [26] Jonathan St BT Evans. Dual-process theories of reasoning: Contemporary issues and developmental applications. *Developmental review*, 31(2-3):86–102, 2011.
- [27] Jonathan St BT Evans and Keith E Stanovich. Dual-process theories of higher cognition: Advancing the debate. *Perspectives on psychological science*, 8(3):223–241, 2013.
- [28] Hans J Eysenck. Dimensions of personality: The biosocial approach to personality. In *Explorations in temperament: International perspectives on theory and measurement*, pages 87–103. Springer, 1991.
- [29] Sebastian Farquhar, Jannik Kossen, Lorenz Kuhn, and Yarin Gal. Detecting hallucinations in large language models using semantic entropy. *Nature*, 630(8017):625–630, 2024.
- [30] Matan Fintz, Margarita Osadchy, and Uri Hertz. Using deep learning to predict human decisions and using cognitive models to explain deep learning models. *Scientific reports*, 12(1):4736, 2022.
- [31] Nigel Ford, David Miller, and Nicola Moss. The role of individual differences in internet searching: An empirical study. *Journal of the American Society for Information Science and technology*, 52(12):1049–1066, 2001.
- [32] Karl Friston. The free-energy principle: a unified brain theory? *Nature reviews neuroscience*, 11(2):127–138, 2010.
- [33] Fiona Fui-Hoon Nah, Ruilin Zheng, Jingyuan Cai, Keng Siau, and Langtao Chen. Generative ai and chatgpt: Applications, challenges, and ai-human collaboration. *Journal of information technology case and application research*, 25(3):277–304, 2023.
- [34] Ulrich Furbach, Steffen Hölldobler, Marco Ragni, and Christian Schön. Bridging the gap: Is logic and automated reasoning a foundation for human reasoning? In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 39, 2017. URL <https://escholarship.org/uc/item/8v89n0dj>.
- [35] Noa Garcia, Chentao Ye, Zihua Liu, Qingtao Hu, Mayu Otani, Chenhui Chu, Yuta Nakashima, and Teruko Mitamura. A dataset and baselines for visual question answering on art, 2020.
- [36] Michael Gerlich. Outsourcing judgment: Hidden anxieties and the rise of cognitive offloading in the age. page 44, 2025.
- [37] Michael Gerlich. Ai tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies*, 15(1):6, 2025.
- [38] Vinod Goel and Raymond J. Dolan. Roles of right prefrontal and lateral temporal cortices in hypothesis generation during reasoning. *NeuroImage*, 26(3):853–861, 2005. doi: 10.1016/j.neuroimage.2005.02.049.

- [39] Olga Golovneva, Moya Peng Chen, Spencer Poff, Martin Corredor, Luke Zettlemoyer, Maryam Fazel-Zarandi, and Asli Celikyilmaz. ROSCOE: A suite of metrics for scoring step-by-step reasoning. In *The Eleventh International Conference on Learning Representations*, 2023. URL <https://openreview.net/forum?id=xYlJRpzZtsY>.
- [40] Samuel D Gosling, Peter J Rentfrow, and William B Swann Jr. A very brief measure of the big-five personality domains. *Journal of Research in personality*, 37(6):504–528, 2003.
- [41] Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, et al. The llama 3 herd of models, 2024. URL <https://arxiv.org/abs/2407.21783>.
- [42] Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Peiyi Wang, Qihao Zhu, Runxin Xu, Ruoyu Zhang, Shirong Ma, Xiao Bi, et al. Deepseek-r1 incentivizes reasoning in llms through reinforcement learning, 2025.
- [43] Gilbert H Harman. The inference to the best explanation. *The philosophical review*, 74(1):88–95, 1965.
- [44] Chaoqun He, Renjie Luo, Yuzhuo Bai, Shengding Hu, Zhen Leng Thai, Junhao Shen, Jinyi Hu, Xu Han, Yujie Huang, Yuxiang Zhang, et al. Olympiadbench: A challenging benchmark for promoting agi with olympiad-level bilingual multimodal scientific problems, 2024.
- [45] Jannica Heinström. Five personality dimensions and their influence on information behaviour, 2003.
- [46] Dan Hendrycks, Collin Burns, Saurav Kadavath, Akul Arora, Steven Basart, Eric Tang, Dawn Song, and Jacob Steinhardt. Measuring mathematical problem solving with the math dataset. *arXiv preprint arXiv:2103.03874*, 2021.
- [47] Jacob B. Hirsh and Jordan B. Peterson. Predicting creativity and academic success with a “fake-proof” measure of the big five. *Journal of Research in Personality*, 42(5):1323–1333, 2008.
- [48] Keith J Holyoak and Robert G Morrison. *The Cambridge handbook of thinking and reasoning*. Cambridge University Press, 2005.
- [49] Pegah Jandaghi, XiangHai Sheng, Xinyi Bai, Jay Pujara, and Hakim Sidahmed. Faithful persona-based conversational dataset generation with large language models, 2023.
- [50] Edwin T Jaynes. Information theory and statistical mechanics. *Physical review*, 106(4):620, 1957.
- [51] Philip N Johnson-Laird. Deductive reasoning. volume 50, pages 109–135. Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA, 1999.
- [52] Philip N Johnson-Laird, Sangeet S Khemlani, and Geoffrey P Goodwin. Logic, probability, and human reasoning. *Trends in cognitive sciences*, 19(4):201–214, 2015.
- [53] Philip Nicholas Johnson-Laird. *How we reason*. Oxford University Press, 2006.
- [54] PN Johnson-Laird, Ruth MJ Byrne, and Sangeet S Khemlani. Models of possibilities instead of logic as the basis of human reasoning. *Minds and Machines*, 34(3):19, 2024.
- [55] John R Josephson and Susan G Josephson. *Abductive inference: Computation, philosophy, technology*. Cambridge University Press, 1996.
- [56] Tom Kibble and Frank H Berkshire. *Classical mechanics*. world scientific publishing company, 2004.
- [57] Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. Large language models are zero-shot reasoners. *Advances in neural information processing systems*, 35:22199–22213, 2022.
- [58] LD Landau and EM Lifshitz. *Mechanics*, volume 1. CUP Archive, 1960.
- [59] X. San Liang and Richard Kleeman. Information transfer between dynamical system components. *Phys. Rev. Lett.*, 95:244101, Dec 2005. doi: 10.1103/PhysRevLett.95.244101. URL <https://link.aps.org/doi/10.1103/PhysRevLett.95.244101>.

- [60] Hunter Lightman, Vineet Kosaraju, Yuri Burda, Harrison Edwards, Bowen Baker, Teddy Lee, Jan Leike, John Schulman, Ilya Sutskever, and Karl Cobbe. Let's verify step by step. 2023.
- [61] Marcia C Linn. *Theoretical and practical significance of formal reasoning*, volume 19. Wiley Online Library, 1982.
- [62] Jian Liu, Leyang Cui, Hanmeng Liu, Dandan Huang, Yile Wang, and Yue Zhang. Logiqa: A challenge dataset for machine reading comprehension with logical reasoning. 2020.
- [63] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.
- [64] Lorenzo Magnani. Model-based and manipulative abduction in science. *Foundations of science*, 9(3):219–247, 2004.
- [65] Maximilian Maier, Vanessa Cheung, and Falk Lieder. Learning from outcomes shapes reliance on moral rules versus cost–benefit reasoning. *Nature Human Behaviour*, pages 1–20, 2025.
- [66] Robert R McCrae and David M Greenberg. Openness to experience. *The Wiley handbook of genius*, pages 222–243, 2014.
- [67] Todor Mihaylov, Peter Clark, Tushar Khot, and Ashish Sabharwal. Can a suit of armor conduct electricity? a new dataset for open book question answering. 2018.
- [68] Ted Moskovitz, Kevin Miller, Maneesh Sahani, and Matthew M Botvinick. A unified theory of dual-process control. 2022.
- [69] Mike Oaksford and Nick Chater. *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford University Press, 2007.
- [70] Art of Problem Solving Foundation. Aime 2024 dataset, 2024. [https://artofproblemsolving.com/wiki/index.php/AIME\\_Problems\\_and\\_Solutions](https://artofproblemsolving.com/wiki/index.php/AIME_Problems_and_Solutions).
- [71] Charles Sanders Peirce. *Collected papers of charles sanders peirce*, volume 5. Harvard University Press, 1934.
- [72] Charles Sanders Peirce. *Collected papers of charles sanders peirce*, volume 5. Harvard University Press, 1934.
- [73] Gordon Pennycook. Chapter three - a framework for understanding reasoning errors: From fake news to climate change and beyond. volume 67 of *Advances in Experimental Social Psychology*, pages 131–208. Academic Press, 2023. doi: 10.1016/bs/aesp.2022.11.003. URL <https://www.sciencedirect.com/science/article/pii/S0065260122000284>.
- [74] Pavlin G Policar and Blaz Zupan. Visualizing highdimensional temporal data using direction-aware t-sne, 2024.
- [75] Shi Qiu, Shaoyang Guo, Zhuo-Yang Song, Yunbo Sun, Zeyu Cai, Jiashen Wei, Tianyu Luo, Yixuan Yin, Haoxu Zhang, Yi Hu, et al. Phybench: Holistic evaluation of physical perception and reasoning in large language models, 2025.
- [76] Iyad Rahwan, Manuel Cebrian, Nick Obradovich, Josh Bongard, Jean-François Bonnefon, Cynthia Breazeal, Jacob W Crandall, Nicholas A Christakis, Iain D Couzin, Matthew O Jackson, et al. Machine behaviour. *Nature*, 568(7753):477–486, 2019.
- [77] Mahdi Ramadan, Cheng Tang, Nicholas Watters, and Mehrdad Jazayeri. Computational basis of hierarchical and counterfactual information processing. *Nature Human Behaviour*, pages 1–15, 2025.
- [78] David Rein, Betty Li Hou, Asa Cooper Stickland, Jackson Petty, Richard Yuanzhe Pang, Julien Dirani, Julian Michael, and Samuel R Bowman. Gpqa: A graduate-level google-proof q&a benchmark. 2024.
- [79] Mohi Reza, Jeb Thomas-Mitchell, Peter Dushniku, Nathan Laundry, Joseph Jay Williams, and Anastasia Kuzminikh. Co-writing with ai, on human terms: Aligning research with user demands across the writing process. *arXiv preprint arXiv:2504.12488*, 2025.
- [80] Lance J Rips. *The psychology of proof: Deductive reasoning in human thinking*. Mit Press, 1994.

- [81] Francisco Salto, Carmen Requena, Paula Alvarez-Merino, Víctor Rodríguez, Jesús Poza, and Roberto Hornero. Electrical analysis of logical complexity: an exploratory eeg study of logically valid/invalid deductive inference. *Brain Informatics*, 10(1):13, 2023.
- [82] Naveen Sangwan. Exploring the big five theory: Unveiling the dynamics and dimensions of personality. volume 1, pages 73–77. 12 2023. doi: 10.60081/SSHA.1.2.2023.73-77.
- [83] Ben Schneiderman. Human-centered ai: ensuring human control while increasing automation. pages 1–2, 2022.
- [84] Keith E. Stanovich and Richard F. West. Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, 23(5):645–665, 2000. doi: 10.1017/S0140525X00003435.
- [85] Keith Stenning and Michiel Van Lambalgen. *Human reasoning and cognitive science*. MIT Press, 2012.
- [86] Jakob Stenseke. On the computational complexity of ethics: Moral tractability for minds and machines. 2023. URL <https://arxiv.org/abs/2302.04218>.
- [87] Gustav Wilhelm Störring. *Experimentelle untersuchungen über einfache schlussprozesse*. W. Engelmann, 1908.
- [88] Alon Talmor, Jonathan Herzig, Nicholas Lourie, and Jonathan Berant. Commonsenseqa: A question answering challenge targeting commonsense knowledge, 2018.
- [89] Qwen Team. Qwen2.5: A party of foundation models, September 2024. URL <https://qwenlm.github.io/blog/qwen2.5/>.
- [90] David Wadden, Shanchuan Lin, Kyle Lo, Lucy Lu Wang, Madeleine van Zuylen, Arman Cohan, and Hannaneh Hajishirzi. Fact or fiction: Verifying scientific claims. 2020.
- [91] Morton Wagman. *Cognitive Science and the Symbolic Operations of Human and Artificial Intelligence: Theory and Research into the Intellective Processes*. Greenwood Publishing Group Inc., USA, 1997. ISBN 0275958531.
- [92] Morton Wagman. *Problem-solving processes in humans and computers: Theory and research in psychology and artificial intelligence*. Praeger Publishers/Greenwood Publishing Group, 2002.
- [93] Shengwei Wang, Keda Chen, Mengduo Yu, Pengjiao Zhao, and Hui Duan. Cdmrnet: multimodal meta-adaptive reasoning network with dynamic causal modeling and co-evolution of quantum states. *Scientific Reports*, 15(1):26370, 2025.
- [94] Zhengnan Xie, Sebastian Thiem, Jaycie Martin, Elizabeth Wainwright, Steven Marmorstein, and Peter Jansen. Worldtree v2: A corpus of science-domain structured explanations and inference patterns supporting multi-hop inference. Proceedings of the Twelfth Language Resources and Evaluation Conference, 2020.
- [95] Chenxu Yang, Qingyi Si, Yongjie Duan, Zheliang Zhu, Chenyu Zhu, Qiaowei Li, Zheng Lin, Li Cao, and Weiping Wang. Dynamic early exit in reasoning models, 2025.
- [96] Zhehao Zhang, Xitao Li, Yan Gao, and Jian-Guang Lou. Crt-qa: A dataset of complex reasoning question answering over tabular data. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 2131–2153, 2023.
- [97] Wanjun Zhong, Ruixiang Cui, Yiduo Guo, Yaobo Liang, Shuai Lu, Yanlin Wang, Amin Saied, Weizhu Chen, and Nan Duan. Agieval: A human-centric benchmark for evaluating foundation models, 2023.
- [98] Zhanke Zhou, Zhaocheng Zhu, Xuan Li, Mikhail Galkin, Xiao Feng, Sanmi Koyejo, Jian Tang, and Bo Han. Landscape of thoughts: Visualizing the reasoning process of large language models. *arXiv preprint arXiv:2503.22165*, 2025.
- [99] Ziyu Zhuang, Qiguang Chen, Longxuan Ma, Mingda Li, Yi Han, Yushan Qian, Haopeng Bai, Zixian Feng, Weinan Zhang, and Ting Liu. Through the lens of core competency: Survey on evaluation of large language models. *arXiv preprint arXiv:2308.07902*, 2023.