Soft Computing
https://doi.org/10.1007/s00500-024-09661-8

FOCUS

Representations of epistemic uncertainty and awareness in data-driven strategies

Mario Angelelli[1] · **Massimiliano Gervasi**[1,2] · **Enrico Ciavolino**[1,3]

Accepted: 9 January 2024

Abstract

The diffusion of AI and big data is reshaping decision-making processes by increasing the amount of information that supports decisions, while reducing direct interaction with data and empirical evidence. This paradigm shift introduces new sources of uncertainty, as limited data observability results in ambiguity and a lack of interpretability. The need for the proper analysis of data-driven strategies motivates the search for new models that can describe this type of bounded access to knowledge.This contribution presents a novel theoretical model for uncertainty in knowledge representation and its transfer mediated by agents. We provide a dynamical description of knowledge states by endowing our model with a structure to compare and combine them. Specifically, an update is represented through combinations, and its explainability is based on its consistency in different dimensional representations. We look at inequivalent knowledge representations in terms of multiplicity of inferences, preference relations, and information measures. Furthermore, we define a formal analogy with two scenarios that illustrate non-classical uncertainty in terms of ambiguity (Ellsberg's model) and reasoning about knowledge mediated by other agents observing data (Wigner's Friend). Finally, we discuss some implications of the proposed model for data-driven strategies, with special attention to reasoning under uncertainty about business value dimensions and the design of measurement tools for their assessment.

Keywords Knowledge representation · Uncertainty modeling · Ambiguity · Data-driven strategy · Big data value · Explainability

1 Introduction

The ongoing technological evolution enables the generation, acquisition, storage, and analysis of an ever-increasing amount of data. In this context, data can be considered raw resources that need to be manipulated, transformed, and combined to extract usable information and knowledge (Ackoff 1989; Ylijoki and Porras 2019). In the literature, data, especially big data, are described in terms of *Velocity*, *Variety*, and *Volume* (Laney 2001) using the original *Vs of the big data* (hereafter referred to as *Vs*). Since 2010, this definition of big data associated with a multi-dimensional characterization has been enriched by new Vs (Hussien 2020). However, some of these features are not intrinsic; namely, they also depend on factors external to the data. Furthermore, their evaluation can be affected by a certain degree of subjectivity. The lack of a context that specifies such characteristics according to the data-driven strategy has occasionally led to confusing definitions (Ylijoki and Porras 2016), making it challenging to identify which Vs to accept (Patgiri and Ahmed 2016). In particular, *Value* is not only listed among the Vs but is also considered a big data feature distinguished from the data features (Uddin and Gupta 2014; Patgiri and Ahmed 2016), as it can be defined according to the other Vs (Geerts and O'Leary 2022) and is more oriented to the way the analyzed data will be used (Ylijoki and Porras 2019). According to these definitions, we refer to *data-driven strategies* as a process of data analysis supported by artificial intelligence (AI)

✉ Mario Angelelli
mario.angelelli@unisalento.it

Massimiliano Gervasi
massimiliano.gervasi@unisalento.it

Enrico Ciavolino
enrico.ciavolino@unisalento.it

[1] CAMPI-Centre for Applied Mathematics and Physics for Industry, University of Salento, 73100 Lecce, Italy

[2] Artificial Intelligence & Data, Deloitte Consulting, S.r.l. S.B., Lecce, Italy

[3] WSB Merito University, Gdansk, Poland

Published online: 29 July 2024

to create *strategic value* in terms of reusable knowledge and decision support (Vitari and Raguseo 2020).

1.1 Motivations

In this context, the focus on big data analysis and the strategic value generated by data-driven strategies should be investigated according to the Knowledge View, as formalized in the Data–Information–Knowledge–Wisdom (DIKW) paradigm (Ackoff 1989). This model has been extended to the Raw Data–Data Formats–Information–Knowledge–Wisdom (RDIKW) paradigm proposed by Wu et al. (2022) and to the Big Data Value Graph, where the RDIKW model is integrated with the Data Mesh (Gervasi et al. 2023a). To support the identification of value generation through data-driven strategies, specific methods have been defined, both in terms of conceptual frameworks and measurement tools. The former encompass value-dimensional frameworks (Elia et al. 2020), while the latter include instruments to assess tangible (e.g., technological) and intangible constructs (e.g., skills or capabilities) that are essential to extracting the potential value of data (Wu et al. 2022). In particular, capabilities associated with agents and technologies can be described in terms of *maturity*, depending on the data-driven strategy. Big data maturity models are widely used by organizations to assess their readiness for adopting and implementing big data-driven strategies (Al-Sai et al. 2022; Vesset et al. 2015; Halper and Krishnan 2013; Corallo et al. 2023.

The extraction of value from big data, often referred to as *potential value*, is generally realized as an estimate of the expected value that an initiative could generate (Ashton 2007; Ishwarappa 2015; Ylijoki and Porras 2016). Specifically, the potential value hidden in (big) data is transformed in the transition from one stage of the DIKW model to another until it becomes visible through the measurement of business performance in the form of business value. In this paper, we focus on knowledge and the sources of uncertainty that could undermine its transition to wisdom.

Actually, most data-driven strategy implementations fail to deliver the estimated value (Reggio and Astesiano 2020), and the causal factors that can explain the failure of data-driven strategies are not contextual and isolated but systemic (Gervasi et al. 2023b). These failures (misalignment between the estimated and generated business value) may come from several factors that characterize data-driven strategies, such as the randomness of the project life cycle, the role of different human or artificial agents, the uncertainty of the results, the multiple characteristics of (big) data, the obsolescence of technologies as a function of time, and the new types of value to be generated (Gervasi et al. 2023b).

This discrepancy between the expected and observed values suggests a need for a deeper investigation to better understand the relation between them. In fact, only a few organizations have implemented a structured value measurement system to rigorously quantify the return on investment in their data-driven strategy (Grover et al. 2018).

Finally, a key point that characterizes data-driven strategies is the role of AI. The limited interpretability of intermediate outcomes during such processes also complicates the adaptation of decision strategies during the initiative. This is now opening up new questions regarding Explainable AI (XAI; see, e.g., Gunning et al. 2019), trustworthy AI (Floridi 2019), as well as human-centered and general-purpose AI.

1.2 Scope of the work

The present work addresses the need for useful representations of uncertainty about knowledge in data-driven strategies. A major source of uncertainty regards the *observability* of data. Specifically, big data are not directly observable; rather, they require a proxy agent (e.g., automated tools) for processing. The effects of this lack of observability on the transition from information to knowledge and wisdom levels in the DIKW hierarchy become particularly evident in analyses based on deep learning, as they are characterized by *black box* approaches that limit the interpretability of extracted information (e.g., selected features). Addressing this issue prompts the search for new approaches that enhance explainability in the adoption of tools relying on AI (Gunning et al. 2019). In this context, our goal is to provide a definition of explainability suitable for the type of uncertainty arising in data-driven strategies, as this uncertainty can undermine the assessment or measurement of the generated value.

For this purpose, we adopt a structural approach for modeling and representing some of the aforementioned notions. Our focus is the (in-)equivalence of relational systems that abstract the notions of accessible knowledge as a resource supporting decision-making and the possibility to transfer (i.e., explain) such knowledge. The objective does not limit itself to knowledge mediated by artificial agents nor to data-driven strategies; in principle, this abstraction level fosters a wider application of the identified properties to different scenarios characterized by "non-classical" (i.e., non-probabilistic) uncertainty. Indeed, we take advantage of formal analogies with other scenarios characterized by ambiguity, such as Ellsberg's urn models (Ellsberg 1961), and explore inequivalent descriptions arising from different data observation levels in measurement settings, such as in Wigner's Friend experiment (Wigner 1995; Frauchiger and Renner 2018). Starting with a dimensional definition of value, we identify order-theoretic structures to evaluate the (lack of) explainability in terms of the update of a knowledge representation.

The relational structure is expressed by endowing dimensional frameworks with enough structure to recognize inconsistencies and ambiguities. In particular, we pay special attention to potential obstructions to the *update* of a knowledge representation in terms of explainability. This can support organizational processes where new evidence and intermediate evaluations during the initiative continuously lead to changing objectives and adopting an agile management methodology.

This high-level framework allows us to identify order-theoretic conditions that enable the description of inconsistencies or bounded resources in the update of a knowledge representation. Order-theoretic notions underlie several decision-making methodologies (Tversky 1969; Jamison and Lau 1973; Greco et al. 2010), which makes the assessment of such conditions scalable in multiple contexts. In the final part of this work, we discuss practical implications and comment on the link with consolidated methodologies to measure abstract constructs (e.g., structural equation modeling, or SEM). This link can prompt future developments in the design of appropriate assessment tools and prevent ambiguities that may undermine a proper assessment of (big) data-driven strategies (Ylijoki and Porras 2016).

The rest of this paper is organized as follows: in Sect. 2, we provide a brief description of uncertainty scenarios that are relevant to our discussion, especially ambiguity, measurements mediated by other agents, and deviations from classical or rational behaviors in decision-making. We start presenting the basic notions of our formalism in Sect. 3, then we pay special attention to the representation of knowledge states as well as their updates and explainability in Sect. 4. The different layers of knowledge inequivalence within our framework, from logical to information-theoretic, are deepened in Sect. 5. In Sect. 6, we formalize the mapping between our formalism and the uncertainty scenarios. In Sect. 7, we discuss our approach in relation to specific forms of uncertainty in measurement models and inequivalence within multi-dimensional frameworks for data-driven strategies. Conclusions and future work are summed up in Sect. 8.

2 Preliminaries on uncertainty scenarios

To better clarify the notion of knowledge within the DIKW hierarchy, we begin by reporting the interpretation of value associated with each stage in Table 1 (Zeleny 1987; Lamba and Dubey 2015).

With reference to this chart, the patterns supporting decision-making should be interpreted or explained in line with a dimensional measurement of the value generated through actions. The occurrence of multiple dimensional assessments may lead to incompatible measurement settings. In the following paragraph, we illustrate relevant sources of indeterminacy that constitute the basis for our model construction.

2.1 Uncertainty types in data-driven strategies

A study by Manyika et al. (2011) predicted that companies using big data in their product and service innovation processes save up to 20−30% on product development costs and achieve faster time-to-market cycles by 50−60%. Similarly, in the public sector, the estimated cost reduction for administrative activities using big data was 15−20%, resulting in a projected generated value of 150–300 billion euros (Cavanillas et al. 2016).

Despite their potential, there is a high failure rate for big data projects, reported to be as high as 85% in Reggio and Astesiano (2020). Montequín et al. (2014) conducted a study identifying and analyzing 26 *failure causes* of Information and Communication Technologies (ICT) projects, as well as 19 *success factors*, using targeted questionnaires. Among the former, the most common causes are incorrect or incomplete definitions of requirements, their continuous change even in the advanced stages of the project, and inaccurate estimations of costs and time. On the other hand, a clear vision of the project objectives and an accurate estimation of feasibility and costs appear among the major success factors.

Here, we concentrate on two types of uncertainty that can have a major impact on the assessment of data-driven strategies. The first type is *ambiguity*, namely the lack of a known or estimated probability distribution over the event space. Ambiguity refers to "*unknown unknowns*", which have been extensively studied in decision- and strategy-making contexts; see, e.g., (Rindova and Courtney 2020). Among the relevant examples of decision frameworks with ambiguity, Ellsberg's urn models are recognized as a paradigm where classical approaches based on maximum expected utility fail to properly represent the preference patterns that are empirically observed (Ellsberg 1961; Sozzo 2020). Ambiguity relates to data-driven strategies mainly due to the lack of accurate or realistic estimates of the raw data value, but also to the mismatch between expected and observed success rates in data-driven strategies.

The definition itself of value is uncertain, as the notion of value is an abstract construct; hence, the measurand is subject to metrological uncertainty. A potential manifestation of such a type of uncertainty is the inconsistency among the multi-dimensional frameworks that arose in the scientific literature to encompass the different realizations of value (Ylijoki and Porras 2016). To reproduce this metrological uncertainty in our formalism, we link to the Wigner's Friend thought experiment (Wigner 1995), which focuses on the implications of non-classical (quantum) measurements. Even though this second scenario originated from the analysis of physical experimental settings, it provides a general basis to

Table 1 DIKW model (Ackoff 1989) associated with the *Taxonomy of knowledge* (Zeleny 1987, Tab. 1) managerial and descriptive interpretation, and the corresponding value associated with each stage (Lamba and Dubey 2015, Tab. 2)

Stage	Taxonomy of knowledge		Value
Data	Muddling through	Know-Nothing	Nothing
Information	Efficiency (measurement+search)	Know-How	Reveals relationships
Knowledge	Effectiveness (decision-making)	Know-What	Reveals pattern
Wisdom	Explicability (judgment)	Know-Why	Reveals action or purpose

illustrate *observer dependence* in evaluations and measurements, which is the second type of uncertainty we focus on in data-driven strategies. In fact, an explicit connection between the Wigner's Friend experiment and frameworks to reason about knowledge (Halpern 2017) was identified (Frauchiger and Renner 2018) and further extended (Nurgalieva and del Rio 2018). More generally, quantum-inspired methods proved useful in modeling the ambiguity in Ellsberg's models (Aerts et al. 2018) and representing deviations from classical behaviors in cognition (Sozzo 2017), decision-making (Sozzo 2020), logic and operational theories (Abramsky and Brandenburger 2014; Abramsky et al. 2017), and social sciences (Cervantes and Dzhafarov 2019).

We introduce these scenarios for ambiguity and observer dependence in evaluations, as we specify the two topics in our framework jointly in Sect. 6.

2.2 Ambiguity: urn models

Urn models encompass a large class of measurement designs that realize various forms of uncertainty, including non-probabilistic ones. A pivotal example is the three-color urn model introduced by Ellsberg (1961, pp. 653–654), which we now briefly describe.

Let us consider an urn containing 90 colored balls, where 30 of them are red and the remaining 60 are yellow or black. The proportion of yellow and black balls is unknown to the decision-maker, who faces cost-free betting alternatives:

1. $\pi_{0,a}$: get 100 if a red ball is drawn from the urn;
2. $\pi_{0,b}$: get 100 if a black ball is drawn from the urn;
3. $\pi_{1,a}$: get 100 if a red *or a yellow* ball is drawn from the urn;
4. $\pi_{1,b}$: get 100 if a black *or a yellow* ball is drawn from the urn.

We introduce the symbol $\prec_{\mathrm{d}}$ to denote the preference relation of a decision-maker between the above-mentioned alternatives. Notably, findings have revealed preferences

$$\pi_{0,b} \prec_{\mathrm{d}} \pi_{0,a} \quad \text{and} \quad \pi_{1,a} \prec_{\mathrm{d}} \pi_{1,b} \tag{1}$$

which contradicts the subjective expected utility theory (Ellsberg 1961). Specifically, (1) violates the *Sure-Thing Principle*, one of the Savage's postulates that can be included in the expected utility theory as a form of monotonicity between preferences and their expected utility. We refer to Aerts et al. (2018) for more details on this topic and for a framework that uses quantum structures to formalize this form of ambiguity in decision-making.

This model is discussed in the context of the framework presented in this work in Sect. 6.1.

2.3 Lack of observability: Wigner's Friend

Wigner's Friend is a central thought experiment in quantum physics, originally conceived by Wigner (see, e.g., Wigner 1995) to highlight the crucial role of observers in quantum measurements. We summarize the key aspects of this thought experiment that are relevant to our scope, directing readers to Frauchiger and Renner (2018) and references therein for a more comprehensive discussion.

We delve into the Wigner's Friend scenario based on the formal analogy between the notion of observability of data in our framework and the role of measurements and observers in quantum physics, both leading to an update of states. Wigner's experiment envisages two laboratories and two observers. The first observer, known as Wigner's Friend, is situated in a laboratory along with a measurement setup. Wigner himself serves as a "super-observer" outside the first laboratory and can perform measurements on it. Wigner's Friend performs the measurement on a physical system (a *spin*), whose possible outcomes are denoted as $|+1\rangle$ and $|-1\rangle$.

The question arises when Wigner's Friend actually observes the outcome of the measurement, while Wigner only knows that the measurement has been performed by his friend, but he has not measured it. For Wigner's Friend, the state is $|+1\rangle$ or $|-1\rangle$, depending on the outcome; on the other hand, Wigner attributes to the combined system in the lab (including the experimental setting and its friend) a superposition of two states, each one associated with the composition (product) of the observed outcome and the state of the friend: the measured system and the friend are *entangled* for Wigner. Then, we have two different perspectives associated with the two observers, which leads to ambiguity about the system's state, namely inconsistency associated with measurements that can be experimentally tested.

2.4 Evidence of non-classical behavior in cognition and self-assessment

The aforementioned scenarios are practical examples that highlight the deviation from classically expected behavior. Similar phenomena are also observed in realistic scenarios, where assumptions to be tested are stated through logical expressions that make use of classical disjunction, conjunction, and negation. These connectives allow for defining events and measuring them using classical approaches rooted in Kolmogorov probability axioms and Bayesian criteria for conditioning and updating probabilities. Measurement tools to assess cognitive constructs, such as questionnaires, enable the estimation of these probabilities from frequencies and the study of empirical correlations. Assumptions regarding abstract constructs and relations among them can be estimated through factor analysis and SEM (Henseler et al. 2015, 2016; Carpita and Ciavolino 2017; Ingusci et al. 2023). In our scope, the constructs of interest are linked to maturity.

Assessing an organization's maturity in terms of capabilities, attitudes, and resources is a key point in defining and implementing data-driven strategies (van de Wetering et al. 2019). Corallo et al. (2023) analyzed the main maturity models according to the three groups of attributes proposed by Mettler et al. (2010). The first group refers to the general attributes, which are inherent in the basic information about models. The second group involves design attributes, which model the structure in terms of evaluation, scope, dimensions, maturity levels, design focus, and evaluation method. Finally, there are attributes related to model application, scope of use (e.g., descriptive, comparative, prescriptive), method of application (e.g., self-assessment, external assessments), and potential availability of supporting material.

Although the models developed and adopted by organizations and analyzed in the literature follow specific standards (Gökalp et al. 2021; de Bruin et al. 2005), the interpretability of the results is complex, and maturity models are often only descriptive or comparative but rarely prescriptive. As highlighted in Corallo et al. (2023), maturity models can have different designs in terms of the number of dimensions and scoring method. Thus, the measured maturity depends on these factors, which affect the measurement's reliability. In addition, it must be considered that the responses collected from respondents may be influenced by biases associated with sample selection. Therefore, the investigation of potential sources of uncertainty in the design of such assessment tools is essential for conducting proper analysis and getting useful insights from the acquired information about capabilities.

Non-classicality arises, for example, when we find a misalignment between syntactic expressions based on classical logic, probability axioms, and empirical frequencies. Incompatibility may correspond to the lack of monotonicity $p(A \wedge B) \leq \min\{p(A), p(B)\}$ for events A, B weighted by the probability $p(\cdot)$. This type of (conjunction) fallacy has been observed in questionnaires (Tentori et al. 2004) and web searches (Sozzo 2017). Even in this case, the explanation of such phenomena can benefit from the Hilbert space representation of quantum mechanics (Sozzo 2020). In Sect. 7, we discuss the implications of our formalism in light of the design of measurement tools that can properly account for non-classical uncertainty in maturity assessment.

Finally, we mention that other deviations from rational (Bayesian) behavior regard contextuality, namely the dependence on an observed property on the whole experimental setting, which includes other simultaneously measured properties. This characterizing aspect of quantum phenomena (Abramsky and Brandenburger 2014; Abramsky et al. 2017) extends to psychological measurements (Dzhafarov and Kujala 2016). Empirical demonstrations of contextuality in psychological assessments have been conducted based on the verification (or violations) of conditions implied by a classical model, namely Bell-type and CHSH inequalities (Cervantes and Dzhafarov 2019).

3 From dimensional frameworks to dimensional structures

Building upon the discussion in the previous sections, we now address the assessment of knowledge value within a decision process for a data-driven initiative. For this purpose, we start with a brief summary of dimensional frameworks in the scientific literature.

3.1 State of the art

Dimensional frameworks are often used to characterize value, enabling its observation and measurement. Grover et al. (2018) distinguished between the *functional value*, e.g., market share and financial return, and the *symbolic value*, which can be identified in the impact on brand and reputation, leading to a positive image as a result of big data analytics (BDA) investment (*signaling effect* or *herding effect*). Günther et al. (2017) conducted a literature review on how organizations realize value from big data through "*paths to value*". In Günther et al. (2017, Sect. 3.1.2), the authors also discussed the current debate regarding the relation between algorithmic and human-based intelligence; we pay attention to this topic in Sect. 4.2. Fosso Wamba et al. (2015) analyzed the five criteria (dimensions) discussed by Manyika et al. (2011) and interpreted them as a different type of generated value.

Gregor et al. (2006) conducted a large-scale survey involving more than a thousand organizations. The collected data also include information regarding ICTs in the organization, the environment, the structure and management practices, and the perceived business value of the use of specific technologies. Finally, factor analysis was used to investigate significant constructs within the high-dimensional survey response data. This approach is of interest in the assessment of the goodness of the chosen dimensions, which can provide insights into possible drivers. The assumptions underlying the choice of model by Gregor et al. (2006) are based on subjective assessments by the authors, which they acknowledge as a limitation of this approach. Second, the results of these analyses contribute to a change in the companies themselves: in the Authors' words, "[*a*] *number of these outcomes equip the firm for further change in a step-by-step process of mutual causation*", which shows how the representation of value can evolve over time according to organizational and contextual changes, and vice versa.

Elia et al. (2020) carried out a systematic literature review to investigate the representations of value; furthermore, they proposed a framework to identify the various types of value, defining 11 *value directions* and grouping them into dimensions. For this purpose, the authors started with the four dimensions of value in Gregor et al. (2006) and considered 22 types of *information technology benefits*. Although the framework presented by Elia et al. (2020) takes its cue from Gregor and co-authors' framework and shares four common dimensions, the two models are distinct and do not lead to the same conclusions: changes prompted by internal (e.g., organizational) or external (e.g., contextual) factors discussed above may require not only the inclusion of a new value dimension but also a different value structure for existing ones.

3.2 Dimensional frameworks as state transitions

We can consider the shift from the model proposed by Gregor et al. (2006) to the one proposed by Elia et al. (2020) as an update of the dimensional architecture representing value. Another case where we can recognize the update of such a dimensional architecture can be found in Maçada et al. (2012). Starting with the models in Gregor et al. (2006) (4 *supra-dimensions*) and Weill and Broadbent (1998) (1 *supra-dimension*), Maçada et al. (2012) identified a new model that confirms the four *supra-dimensions* in Gregor et al. (2006) but permutes the *sub-dimensions*. This aspect is worth considering because it stresses a subjective component in dimensional definitions, which are representative of a particular view, not only in terms of granularity but also in terms of classification.

We start introducing our conceptual formalism to encompass this type of update by specifying the role of dimensional frameworks within an assessment process. Choosing a set of evaluation stages within the time span of the process, referred to as *process states*, we focus on classes of transitions between them to highlight the relational aspects of value. For each pair of states ψ_1 and ψ_2, we associate a labeled transition between them and denote it as $\psi_1 \xrightarrow{\tau} \psi_2$. In this way, given a state ψ_1, the class of all the possible transitions τ originating from ψ_1 defines the possible inferences that an agent can make starting from ψ_1.

Now, we include in our model the occurrence of multiple dimensions that guide the decision process and its evaluation. The minimal structure that we assume to support decision-making criteria is an order relation. So, we provide the following:

Definition 1 Let $\mathcal{V}$ be a set of non-empty partially ordered sets (*posets*; see, e.g., Davey and Priestley (2002)). We label each element of $\mathcal{V}$ through a totally ordered set $\mathcal{I}$, so we can express

$$\mathcal{V} := \{(V_i, \preceq_i) : i \in \mathcal{I}\}, \tag{2}$$

where $\preceq_i$ is a reflexive, antisymmetric, and transitive relation on V_i for each $i \in \mathcal{I}$.

Let $\mathcal{J} \subseteq \mathcal{I}$. We consider the categorial product of the latent dimensions $V_i, i \in \mathcal{J}$, which is the well-known Cartesian product for the category of sets (**Set**):

$$\varrho_{\mathcal{J}} := \prod_{j \in \mathcal{J}} V_j, \tag{3}$$

where the order of factors is derived from the order in $\mathcal{I}$. The corresponding *value frame* is then defined as the disjoint union of such products for all non-empty subsets of $\mathcal{I}$:

$$\varrho := \coprod_{\emptyset \subset \mathcal{J} \subseteq \mathcal{I}} \varrho_{\mathcal{J}}, \tag{4}$$

where $\coprod$ denotes the disjoint union (or coproduct) of sets.

For the sake of concreteness, we present an example with relevant dimensions for data-driven strategies by recalling the framework provided by Elia et al. (2020).

Example 1 The definition of the dimensions in Elia et al. (2020, Tab. 10) refers to big data. In this example, we report two selected dimensions $V_1 = \{m_1, m_2, m_3, m_4, m_5\}$ and $V_2 = \{b_1, b_2, b_3\}$, where the interpretation of these labels is detailed as follows (Elia et al. 2020, Tab. 5):

- V_1: *Strategic value*
  1. m_1: "*New competitive advantage*"
  2. m_2: "*Alignment between IT and business strategy*"
  3. m_3: "*Quicker response to change*"
  4. m_4: "*More effective customer relationships*"
  5. m_5: "*Better products and services*"

- V_2: *Transformational value*
  1. b_1: "*Reinforcement of organizational capabilities*"
  2. b_2: "*Innovation in business models*"
  3. b_3: "*Efficiency in organizational structure and processes*".

The authors drew these dimensions from Gregor et al. (2006), who used factor analysis to investigate survey responses and associate 22 benefit items with 4 dimensions. In particular, "*Expanding organizational capabilities*" has factor loadings of 0.49 and 0.44 along the Transformational and Strategic benefits, respectively.

Starting with dimensions $V_1, \ldots, V_n$, we can construct a new representation by introducing $\mathcal{I} := \wp(\{1, \ldots, n\}) \backslash \{\emptyset\}$, the set of non-empty subsets of $\{1, \ldots, n\}$, and dimensions $W_{\mathcal{S}}$, $\emptyset \subset \mathcal{S} \subseteq \{1, \ldots, n\}$. The elements of $W_{\mathcal{S}}$ are constructed from the values of benefit items in $\bigcap_{j \in \mathcal{S}} V_j$ when this intersection is not empty. This allows distinguishing the membership of the "*Expanding organizational capabilities*" item in different dimensions based on distinct interpretations.

If one relies solely on the magnitude of factor loadings as a membership criterion to specify the dimensional framework, then the narrow difference between the loadings in the previous example might suggest an association of the benefit "*Expanding organizational capabilities*" with both the Transformational and Strategic and dimensions. This points out the need to assess the identity of variables among distinct dimensions. Each individual study can adopt other quantitative criteria to confirm the discriminant validity of the constructs and dimensions in a structural model (Henseler et al. 2015). However, the combination of multiple studies should assess whether the measurement models are compatible. This fundamental requirement is made explicit for individual studies involving a multi-group analysis (Henseler et al. 2016; Ingusci et al. 2023), where the configural invariance is the first step to be checked to make this comparison meaningful (Henseler et al. 2016, pp. 413–414).

The comparability of distinct frameworks leads to additional issues regarding the identity of the dimensions investigated in the individual studies, even in terms of the interpretation of the indicators within the different measurement settings. Products and disjoint unions in (4) can represent such deviations from unidimensionality in the sense of multiple latent dimensions related to a group of indicators. In data-driven strategies, such a distinction acquires more relevance due to the intrinsic ambiguity in the definition of (big) data-characterizing features (Ylijoki and Porras 2016).

However, value frames in Definition 1 do not take into proper account the lack of knowledge regarding the whole class of dimensions and potential interdependencies among them. Therefore, along with products and disjoint unions, we should also include meta-reasoning to highlight inconsistencies in the comparison of dimensional structures. For example, we can encode knowledge about the framework dimensionality through the inclusion of a representative of the state itself, e.g., a dimension $V_{\text{dim}} := \wp(\mathbb{N}_0)$ that addresses the number of potential dimensions of the knowledge state. The choice of $\kappa \in \varrho_{\mathcal{J}}$ lets us represent uncertainty about the dimensionality ($\dim \in \mathcal{J}$ and $\#\kappa(\dim) > 1$, where $\#S$ is the cardinality of a set S), as well as the lack of specification ($\dim \notin \mathcal{J}$). Furthermore, inconsistency arises when $\#\mathcal{J} \notin \kappa(\dim)$.

Next, we extend the value frame ϱ by endowing it with relational structures and conditions to formalize the types of uncertainty mentioned in Sect. 2.1.

4 Model proposal: operational aspects of knowledge representations

In this section, we explore the conditions that make knowledge transferable between different representations and, hence, reusable.

4.1 Comparison, composition, and the role of observers: inner states

The assumption that each dimension V_i is endowed with a partial order $\preceq_i$, i.e., a reflexive, symmetric, and transitive, but not necessarily total relation, is in line with well-established methods to formalize concept analysis and knowledge structures (Doignon and Falmagne 2012). A stronger assumption that can be considered is the existence of an associative operation $\oplus_i$ for each latent dimension V_i that is *idempotent*, i.e., $x \oplus_i x = x$ for all $x \in V_i$ and $i \in \mathcal{I}$. This operation defines an order relation as follows:

$$\forall a, b \in V_i : \quad a \preceq_i b \Leftrightarrow a \oplus_i b = b. \tag{5}$$

Such an operation, which combines the partial order $\preceq_i$ and the notion of composition, is essential for the subsequent analysis of knowledge representations in the context of data-driven strategies. To establish an algebraic structure for comparing and composing different knowledge states, we endow the value frame ϱ with an order relation $\preceq$ as follows: we express the content of each individual element $\kappa \in \varrho$ as a partial function from $\mathcal{I}$ to $\coprod_{i \in \mathcal{I}} V_i$ by defining, for a given $\mathcal{J} \subseteq \mathcal{I}$, the map

$$\kappa : \mathcal{J} \longrightarrow \bigcup_{i \in \mathcal{I}} V_i, \quad j \mapsto \kappa(j) \in V_j. \tag{6}$$

For each $\mathcal{H} \subseteq \mathcal{J} \subseteq \mathcal{I}$, $\pi_{\mathcal{H}}$ denotes the projection of $\varrho_{\mathcal{J}}$ onto $\prod_{h \in \mathcal{H}} V_h$, and we explicitly write $\pi_{\mathcal{J},i}$ to refer to the canonical projection onto V_i for each $i \in \mathcal{J}$. Let us denote the domain of such a partial function as $\underline{\kappa}$. For all κ_1, κ_2 with

$\underline{\kappa_1} = \underline{\kappa_2}$, we consider $\kappa_1 \preceq \kappa_2$ if

$$\forall i \in \underline{\kappa_1} : \pi_{\underline{\kappa_1},i}(\kappa_1) \preceq_i \pi_{\underline{\kappa_2},i}(\kappa_2). \tag{7}$$

This definition is in line with the notion of product category, with particular reference to posetal categories in the present discussion. Then, we extend this order by taking into account the domain of the elements of ϱ:

$$\kappa_1 \preceq \kappa_2 \Leftrightarrow \underline{\kappa_1} \subseteq \underline{\kappa_2} \quad \text{and} \quad \kappa_1 \preceq \kappa_2|_{\underline{\kappa_1}}, \tag{8}$$

where $\kappa_2|_{\underline{\kappa_1}}$ denotes the restriction of κ_2 to the domain $\underline{\kappa_1}$ of κ_1. We denote the resulting poset $(\varrho, \preceq)$ as $\varrho^\star$. Now, we can state the following:

Definition 2 An *inner state* κ is an element of the poset $\varrho^\star$. With a slight abuse of notation, we can equivalently express each individual inner state κ as a partial function with domain $\underline{\kappa}$, in line with (6). In this way, we can compare inner states by domain extension (8) and componentwise ordering (7).

The statement $a \preceq_i b$ can be interpreted as follows: let us consider any two agents $\mathtt{A}, \mathtt{B}$ with inner states $\kappa_\mathtt{A}, \kappa_\mathtt{B}$, respectively, such that $i \in \underline{\kappa_\mathtt{A}} \cap \underline{\kappa_\mathtt{B}}$, $a = \pi_{\underline{\kappa_\mathtt{A}},i}(\kappa_\mathtt{A})$, and $b = \pi_{\underline{\kappa_\mathtt{B}},i}(\kappa_\mathtt{B})$. Then, all the knowledge value recognized by Agent $\mathtt{A}$ along the dimension V_i is also recognized by Agent $\mathtt{B}$. In our perspective, the order relation represents the possible inferences that can be drawn by an agent using its knowledge resources.

Remark 1 The focus on the domain of an inner state is of major relevance in defining awareness in the present formulation of data-driven strategies. Indeed, the statement $i \in \underline{\kappa}$ is interpreted as the assertion that the agent can evaluate its knowledge along the dimension V_i. If $\perp_i$ is the minimum element of V_i, assuming it exists, the statement $\kappa(i) = \perp_i$ means that the agent knows she does not know about the dimension V_i. On the contrary, $i \notin \underline{\kappa}$ means that the agent is unaware of the dimension V_i.

This can be read in relation to the operator K_i associated with Agent i in modal logic; the combination of the notions that are used to model knowledge structures (in particular, Kripke structures) in the Wigner's Friend extended scenarios designed by Frauchiger and Renner (2018) is discussed in Nurgalieva and del Rio (2018).

The order (8) extends (7) by relaxing a comparability condition, from $\underline{\kappa_1} = \underline{\kappa_2}$ to $\underline{\kappa_1} \subseteq \underline{\kappa_2}$. An analogous distinction was pointed out in Angelelli (2017, Sect. 3) to examine inequivalent representations in a statistical physical context. In the scope of this work, we note the following:

Remark 2 The domain extension (8) establishes a connection between different dimensional contexts $\varrho_\mathcal{J}$. This generates the order compatibility:

$$\kappa_1|_{\underline{\kappa_1} \cap \underline{\kappa_2}} \preceq \kappa_2|_{\underline{\kappa_1} \cap \underline{\kappa_2}} \Rightarrow \kappa_1|_{\underline{\kappa_1} \cap \underline{\kappa_2}} \preceq \kappa_2. \tag{9}$$

The order (8) is a *value-based* view, as it does not differentiate between different domains due to (9). Other orders can be associated with the set ϱ; in particular, we can compare two elements only if they have the same domain:

$$\kappa_1 \sqsubseteq \kappa_2 \Leftrightarrow \underline{\kappa_1} = \underline{\kappa_2} \text{ and } \forall i \in \underline{\kappa_1} : \kappa_1(i) \preceq_i \kappa_2(i). \tag{10}$$

This second order is a *domain-based* view, leveraging the representation of inner states as partial functions. In this sense, the comparability of two partial functions is not based only on the value of potential input dimensions but also on the fact that they are functions with the same domain. This defines a new poset, $\varrho_\star := (\varrho, \sqsubseteq)$. Clearly, $\varrho^\star$ is an extension of $\varrho_\star$ since each pair $\kappa_1 \sqsubseteq \kappa_2$ corresponds to a pair $\kappa_1 \preceq \kappa_2$ in $\varrho^\star$.

The two posets coincide only under the unidimensionality condition $\#\mathcal{I} = 1$. In the multi-dimensional case, we refer to $\varrho_\star$ and $\varrho^\star$ as *lower* and *upper* posets, respectively. This terminology adapts the lower and upper probabilities (or belief and plausibility, respectively) that are used to model imprecise probability, e.g., in Dempster–Shafer theory (Halpern 2017, Sect. 2.3–2.4) [also see Cuzzolin (2020) for a geometric view of these notions].

4.2 Self-reference

The distinguishing role of the domain can be used as a basis to consider different orders starting from the same class of posets. In turn, this allows for modeling the partiality of the composition (5). This objective fits the scope of this work, as the (lack of) composition of knowledge states can be linked to the (lack of) explainability. For example, given the knowledge representations $\kappa_\mathtt{H}$ and $\kappa_\mathtt{A}$ of a human agent $\mathtt{H}$ and an artificial agent $\mathtt{A}$, respectively, the composition $\kappa_\mathtt{A} \oplus \kappa_\mathtt{H}$ may not be feasible when the knowledge of $\mathtt{A}$'s inner state $\kappa_\mathtt{A}$ is partially accessible to $\mathtt{H}$.

Looking at the scenarios described in Sect. 2, the type of uncertainty we want to describe primarily arises through meta-reasoning, which, in our context, involves the composition or comparison of inner states. In particular, the ambiguity in Ellsberg's urn model is not manifest when considering individual decision contexts (bets $(\pi_{0,a}, \pi_{0,b})$ and $(\pi_{1,a}, \pi_{1,b})$), but it emerges only when both of these contexts are taken into account and compared. Similarly, the Wigner's Friend phenomenon involves Wigner's reasoning about its friend's state, as elaborated in Frauchiger and Renner (2018).

Then, we can use the two posets $\varrho_\star$ and $\varrho^\star$ defined in Sect. 4.1 to represent this form of meta-reasoning.

Definition 3 We extend $\varrho^\star$ by appending the poset $(\varrho, \sqsubseteq)$ as a new dimension and repeating the construction in (7)–(8). The corresponding poset

$$\mathrm{K} := (\varrho \sqcup \varrho \sqcup (\varrho \sqcap \varrho), \preceq_{\mathrm{K}}) \tag{11}$$

with the order $\preceq_{\mathrm{K}}$ induced by the aforementioned construction is referred to as the poset of *lower meta-states*.

In Remark 2, $\varrho_\star$ is introduced as a lower poset, namely, the inclusion $\sqsubseteq \subseteq \preceq$ between relations on ϱ holds. In analogy to lower and upper posets, we can also consider a different definition of meta-states where $\varrho^\star$ plays the role of the lower poset, and we extend the order $\preceq$ in the new dimension. In particular, we can consider a linear extension $\leq$ of $\preceq$ using a monotone function $h : \varrho^\star \longrightarrow \mathbb{R}$:

$$\kappa_1 \preceq \kappa_2 \Leftrightarrow h(\kappa_1) \leq h(\kappa_2). \tag{12}$$

We use this representation in the following sections to provide an information-theoretic view of the state, where the extension $\preceq \subseteq \leq$ is derived from an information measure. This type of meta-reasoning can be expressed in a more general setting by introducing a second type of meta-states, which rely on a poset V_{exp} that expresses knowledge about the relational structure in $\varrho^\star$.

Definition 4 Let us consider a poset V_{exp} and an order-preserving mapping $\nu : \varrho^\star \longrightarrow V_{\mathrm{exp}}$. Then, we append V_{exp} as a new dimension and obtain the poset

$$\mathrm{K}^{(\nu)} := \left(\varrho \sqcup V_{\mathrm{exp}} \sqcup (\varrho \sqcap V_{\mathrm{exp}}), \preceq_\nu\right), \tag{13}$$

where $\preceq_\nu$ in (13) is obtained through the construction in (7)–(8). The elements of the poset $\mathrm{K}^{(\nu)}$ are referred to as *ν-meta-states*.

4.3 Explainability as compositional existence

Finally, we use the notions introduced above to provide a formal definition of explainability in our context.

Definition 5 A lower meta-state is called *diagonal* if it can be expressed as $(\kappa, \kappa) \in \mathrm{K}$ for some inner state $\kappa \in \varrho^\star$. The *update* of $\kappa \in \varrho^\star$ by lower meta-states is the composition $\kappa \vee \psi$ with another state $\psi \in \varrho^\star$, when the supremum exists. Such an update is said to be *explainable* when $(\kappa, \kappa) \vee (\psi, \psi)$ also exists in K and, hence, is diagonal.

Similarly, a ν−meta-state is diagonal if it has the form $(\kappa, \nu(\kappa))$ for some $\kappa \in \varrho^\star$. An explainable update by ν-meta-states is a composition (supremum) of diagonal elements in $\mathrm{K}^{(\nu)}$ that is diagonal too.

This definition specifies the possibility of extending an inner state and distinguishing extensions that change the knowledge base and, hence, are inconsistent with respect to the current dimensional setting. Note that the focus of explainability in this context is on knowledge updates, in agreement with the attention paid to (non-)reusable knowledge in data-driven strategies. When a combination $K \vee \kappa_{\mathbb{A}}$ is not feasible or is not compatible with the explainability accessible to K, the two inner states cannot be combined into a new state of knowledge. In turn, this may impede the reuse of the knowledge resource carried by another inner state in other contexts. An obstruction to the existence of such a state is the presence of multiple, non-equivalent value representations that an agent cannot directly discern. This is the situation we want to explore to describe interactions between agents in structural terms.

Remark 3 The previous definition stresses the role of explainability in relation to accessible knowledge through the projections $\pi_{\underline{\kappa},i}$ (see Definition 2). According to Remark 2, the role of the domain of inner states in the definition of K, e.g., through the lower poset $\varrho_\star$ in (10), relates the existence of an explanation for a lower meta-state to the existence of the same set of projections $\pi_{\mathcal{J}}$, $\mathcal{J} \subseteq \underline{\kappa}$, for both κ and $\kappa \vee \psi$.

In the following sections, we analyze the definitions provided above through inconsistencies that cannot be resolved through an explainable update of an inner state. Within a data-driven initiative, this allows assessing whether data and other agents that can observe them generate *reusable* knowledge (an explainable update) or not.

5 Uncertainty and inequivalence in knowledge representations

The labeling induced by the disjoint product (4) in this framework is analogous to indexation-by-conditions in the contextuality-by-default approach in cognitive sciences, where variables are indexed by the context they are part of Dzhafarov and Kujala (2016). To remove the dependence of such context, we can represent each element of $\kappa \in \varrho^\star$ as a tuple of pairs $(i, \kappa(i))$; then, we consider the equivalence relation defined by the projection onto the second coordinate, so $(i, \kappa(i))$ is identified with $(j, \kappa(j))$ if and only if $\kappa(i) = \kappa(j)$.

Such labelings generate a connection between different posets, as elements of $V_i \cap V_j$, $i, j, \in \underline{\kappa}$, act as a linkage between the underlying posets V_i and V_j. We can formalize this linkage by extending the focus from equal to corresponding elements through order-preserving mappings. This extension lets us consider the extent of order compatibility between different dimensions, since order-preserving mappings may not preserve compositions (suprema). Given the interpretation of explainable updates based on compositions

(Definition 5), these order-preserved mappings can be used to realize inequivalent knowledge representations.

Next, we present three instances of knowledge inequivalence from the logical, order-theoretic, and information-theoretic perspectives, respectively.

5.1 Non-classicality from multiple potential implications

The DIKW hierarchy mentioned in the Introduction distinguishes access to data from the decision-making stages, which can be summarized as a set of inferences to draw conclusions about the effect of actions based on evidence and empirical premises.

The first layer where the inequivalence of knowledge representations can be addressed is the logical one, where inference is expressed as material implication. Specifically, uncertainty emerges as the occurrence of multiple inequivalent implications, which can be obtained in our context through the following construction.

Example 2 Let us consider a set-theoretic representation of a finite distributive lattice as an appropriate subset of the power set $\wp(\mathcal{S})$, which is always possible due to Birkhoff's representation theorem (Davey and Priestley 2002, Sect. 5.12). It is well known that the implication $A \rightarrow \cdot$ defined by $A \rightarrow Y := Y \cup (A)^{\complement}$ is the upper adjoint of the conjunction $A \cap \cdot$; namely they are monotonic functions satisfying the relation

$$A \cap Y \subseteq Z \Leftrightarrow Y \subseteq A \rightarrow Z, \quad A, Y, Z, \subseteq \mathcal{S}, \tag{14}$$

where $\complement$ denotes the set-theoretic complement with respect to $\mathcal{S}$. The implication is a fundamental logical connective to describe inference, and the adjointness condition is the basis for generalizing classical logic to Heyting algebras and extended-order algebras in the context of fuzzy operators (see, e.g., Della Stella and Guido 2012 and references therein).

However, this construction presumes the knowledge of the whole set $\mathcal{S}$ to evaluate the complement $\complement$. In particular, we can model partial knowledge on $\mathcal{S}$ by considering a class $\{\mathcal{S}_u,\ u \in \{1, \dots, n\}\}$ of n potential spaces that define as many implications $\rightarrow^u$, $u \in \{1, \dots, n\}$.

Remark 4 From the previous argument, we can see that the uncertainty about the base set $\mathcal{S}$ entails a deviation from classicality. Indeed, we have two alternatives for a finite poset K. When K is a distributive lattice, the previous example shows that the lack of knowledge of the full set of dimensions generates multiple implications. Otherwise, the poset is not distributive, which is a main deviation from classical logic that is used to characterize quantum logics (in particular, by replacing distributivity with modularity; see, e.g., Harding 1996) and their extensions. In both alternatives, we infer a non-classical behavior from bounded knowledge resources.

From the logical layer, we can move our focus to the accessibility of potential inferences that can be drawn. In turn, this interpretation exploits the order structure of the dimensions and entails a second layer of inequivalence in knowledge representations.

5.2 Inequivalent representations from accessibility boundary

Inequivalent descriptions of a statistical system are often a hint of its non-trivial characteristics. For instance, in Angelelli (2017, Sect. 3), a limiting procedure changes the compositional structure of the statistical model, returning an instance of (5) with connections to fuzzy sets. Such a limit accounts for a parameter configuration suited to exponential degenerations of energy levels. In view of our formalization of the Wigner's Friend scenario, we investigate this aspect in our setting through the following example.

Example 3 Let us take a class of dimensions $\mathcal{V}$ and a distinguished dimension V_i where the order relation $\preceq_i$ is partial. Furthermore, suppose that the supremum $a \vee_i b$ exists for all $a, b \in V_i$, which is an instance of (5). Another way to encode the relation $\preceq_i$ is to introduce

$$\begin{aligned} x \in V_i \mapsto \iota_i(x) &:= \{y \prec_i x : y \in V_i\}, \\ (V_i^\star, \preceq_i^\star) &:= (\{\iota_i(x) : x \in V_i\}, \subseteq), \end{aligned} \tag{15}$$

where $a \prec_i b$ means $a \preceq_i b$ and $a \neq b$. This representation of value is not based on a single valuation $x \in V_i$ but on the inferences that can be drawn through non-trivial processing of the available information encoded in x.

The statement $a \preceq_i b$ implies that $\iota_i(a) \subseteq \iota_i(b)$, so ι is a strictly monotone mapping; in this sense, the order relation between compatible elements is preserved passing from V_i to $V_i^\star$ (i.e., an order morphism). However, these two representations differ when the composition structure (5) is considered: we can consider $\vee$ (the supremum operation) and $\cup$ as two operations satisfying (5) for V_i and $V_i^\star$, respectively. According to Angelelli (2017, Prop. 5.1), we know that $(V_i, \vee)$ is homomorphic to $(V_i^\star, \cup)$ only if $(V_i, \preceq_i)$ is totally ordered. Since the order relation attributed to V_i is not total, these different definitions of the dimension produce inequivalent results under composition.

The final layer is the quantitative one, which relies on measures to evaluate the amount of knowledge encoded in a given representation. In the following subsection, we adapt the notion of information measure to our knowledge-based view.

5.3 Uncertainty in information measures

The lack of knowledge about the entire dimensional space ($\mathcal{S}$ in Sect. 5.1, unique maximal elements of $\iota(x)$ in Sect. 5.2) affects the quantification of the information content in knowledge representations. This lets us establish the analogy with the role of normalization of subsystems discussed in Angelelli (2017, Sect. 7). Consider a state κ and the ordered set of non-negative reals $(\mathbb{R}_{\geq 0}, \leq)$. Associate each dimension V_i, $i \in \underline{\kappa}$, with a weight $w_\kappa(i)$ for a given function $w_\kappa : \underline{\kappa} \to \mathbb{R}_{\geq 0}$. In particular, we can take a monotone function $\nu : \varrho^\star \longrightarrow \mathbb{R}_{\geq 0}$, which was also mentioned for the construction of ν-meta-states in (13), and obtain w_κ via $w_\kappa(i) := \nu\left(\kappa|_{\{i\}}\right)$. In this way, we take into account both the domain and the values of κ; by focusing on the dimension $\varrho_\star$ introduced in (10), we can also associate each $\kappa \in \varrho_\star$ with a weight $w_\kappa(\varrho_\star) \geq 0$, so that $\kappa_1 \sqsubseteq \kappa_2$ implies $w_{\kappa_1}(\varrho_\star) \leq w_{\kappa_2}(\varrho_\star)$. On the other hand, the attribution $W_\kappa(i) := \nu\left(\bigvee V_i\right)$ only depends on the supremum $\bigvee V_i$ of V_i and provides us with a definition of normalization suited to our context, which only relies on the domain $\underline{\kappa}$.

Then, to identify uncertainty within information measures, we can use a representation where each $V_i \in \mathcal{V}$ is isomorphic to $(\mathbb{R}_{\geq 0}, \leq)$. In this setup, each $\kappa \in \varrho$ induces, after normalization, a probability distribution whose support is $\underline{\kappa}$. We consider as an information measure $h : \varrho \longrightarrow \mathbb{R}_{\geq 0}$ the normalized Shannon entropy (Halpern 2017, Sect. 3.11)

$$h(\kappa) := \frac{H(\kappa)}{H_{\max}(\underline{\kappa})} = -\frac{1}{\ln\left(\#(\underline{\kappa})\right)} \cdot \sum_{t \in \underline{\kappa}} \frac{w_t}{\sum_{u \in \underline{\kappa}} w_u} \cdot \ln\left(\frac{w_t}{\sum_{u \in \underline{\kappa}} w_u}\right). \tag{16}$$

This function is widely adopted to quantify the uncertainty (or, dually, the information and complexity) within a given distribution. The inclusion of the normalization $\ln\left(\#(\underline{\kappa})\right)^{-1}$ derived from the maximum entropy achievable for a distribution with support $\underline{\kappa}$ takes into account the *potential* probability assignments to the available dimensions. Such a dependence makes the normalized entropy undefined, and, when available, partial information about the support $\underline{\kappa}$ in terms of lower or upper approximations induces bounds for $\ln(\#(\kappa))^{-1}$.

In this setting, each poset $\varrho_\star$ and $\varrho^\star$ entails order conditions between probability distributions. This kind of comparison also arises in information theory when one considers the *Kullback–Leibler divergence* (Halpern 2017, Ch. 3) to quantify the differences from the update of a probability distribution. Denoting as $p(\kappa_1)$ and $p(\kappa_2)$ the probability distributions associated with κ_1 and κ_2, respectively, the Kullback–Leibler divergence $D_{\mathrm{KL}}(p(\kappa_2)||p(\kappa_1))$ from $p(\kappa_1)$ to $p(\kappa_2)$ can be evaluated only if $p(\kappa_2) = 0$ whenever $p(\kappa_1) = 0$. This assumption of absolute continuity formalizes the constraint that the support of the distributions (interpreted as the set of elements with positive probability weight) does not increase. In our framework, we include the possibility to extend the support $\underline{\kappa_1}$ with new dimensions; on the other hand, this extension is evaluated differently in the two posets $\varrho_\star$ and $\varrho^\star$, which gives a criterion to discriminate explainable updates.

6 Modeling in data–agent interactions

Now, we examine the uncertainty scenarios described in Sect. 2 within the proposed framework.

6.1 Ambiguity and data–agent interactions

6.1.1 Preliminary discussion

Before defining the connection between Ellsberg's three-color urn model and our formalism in Sect. 6.1.2, we identify some preliminary analogies to contextualize the decision-making problem in the scope of this work. The decision-maker is represented by a human agent $\mathtt{H}$, who has access to information regarding the value of the "Red" dimension; specifically, $\mathtt{H}$ can assess the risk associated with "Red", e.g., knowing its impact and probability. On the contrary, the information possessed by $\mathtt{H}$ about the remaining two colors, "Black–Yellow", only acknowledges their existence and their cumulative probability weight ($\frac{2}{3}$).

A second agent, which we can associate with artificial intelligence (AI) with an inner state denoted as $\kappa_{\mathtt{A}}$, can get access to data to recognize the value of the "Black–Yellow" information dimension to a greater extent with respect to $\mathtt{H}$. In particular, there may be a latent factor in the data that lets $\mathtt{A}$ distinguish two "Black" and "Yellow" information dimensions. The human agent knows that $\mathtt{A}$ is able to recognize new value in the "Black–Yellow" dimension.

Ellsberg's paradox corresponds to the misalignment between these information dimensions and knowledge dimensions, namely, the two decision contexts corresponding to the two lotteries $(\pi_{0,a}, \pi_{0,b})$ and $(\pi_{1,a}, \pi_{1,b})$. The existence of $\mathtt{A}$ allows the extraction of value from the "Black-Yellow" information dimension, but this cannot prompt a change of knowledge state for $\mathtt{H}$ that is able to discern the value of the "Black" and the "Yellow" evaluations.

We provide a diagrammatic depiction of this phenomenon in Fig. 1.

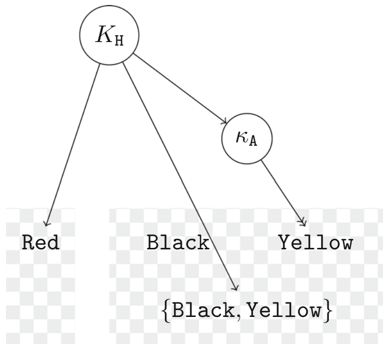


Fig. 1 Diagrammatic representation of Ellsberg's three-color model compared to the relation of human and artificial agents with data

6.1.2 Ambiguity and explainability of knowledge updates

Now, we provide a formal correspondence between Ellsberg's three-color model and the present framework. We introduce the notation $\mathtt{B}^{(n)}$ for the Boolean algebra $\left(\wp(\{1,\ldots,n\}), \cup, \cap, \cdot^{\mathrm{C}}, \{1,\ldots,n\}, \emptyset\right)$ of the power set $\wp(\{1,\ldots,n\})$. Consider two dimensions, $\mathtt{B}_0$ and $\mathtt{B}_1$, isomorphic to $\mathtt{B}^{(1)}$. These two objects abstract the two different observation/measurement settings, i.e., the two lotteries $(\pi_{0,a}, \pi_{0,b})$ and $(\pi_{1,a}, \pi_{1,b})$ in Ellsberg's model. The set K is in bijection with

$$\wp(\{\top^{(0)}\}) \sqcup \wp(\{\top^{(1)}\}) \sqcup \wp(\{\top^{(0)}, \top^{(1)}\}), \tag{17}$$

where we use the labelings (0) and (1) to distinguish the two lotteries as a result of the disjoint union. Then we focus on the explainability of the composition $\{\top^{(0)}\} \vee \{\top^{(1)}\}$ of $\{\top^{(0)}\} \in \mathtt{B}_0$ and $\{\top^{(1)}\} \in \mathtt{B}_1$, which represents an update where an agent becomes aware of a second decision scenario (lottery). We obtain

$$\{\top^{(0)}\} \vee \{\top^{(1)}\} = \left(\{\top^{(0)}\}, \{\top^{(1)}\}\right) \quad \text{in } \varrho^{\star}. \tag{18}$$

On the other hand, this composition is not defined in $\varrho_{\star}$.

An analogous result is obtained using ν-meta-states as in Definition 5. Here, we consider $\nu : \varrho^{\star} \longrightarrow \mathtt{B}^{(1)}$ with $\nu(\kappa) = \{1\}$ at $\underline{\kappa} = \{1, 2\}$, and $\nu(\kappa) = \emptyset$ otherwise. From (18), we find

$$\nu(\{\top^{(0)}\}) = \nu(\{\top^{(1)}\}) = \emptyset \subset \{1\} = \nu\left(\{\top^{(0)}\} \vee \{\top^{(1)}\}\right), \tag{19}$$

so the composition of the two lotteries is not explainable. In Fig. 2a, b, we provide a graphical representation of the previous argument based on Hasse diagrams (Davey and Priestley 2002, Sect. 1.15), denoting $\ell_0 := \{\top^{(0)}\}$ and $\ell_1 := \{\top^{(1)}\}$ to stress the link to the two lotteries in Ellsberg's model.

We point out that a different view on an analogous phenomenon was given in Angelelli (2017, Sect. 7), where different orders for quantities characterizing physical subsystems emerge as a consequence of different choices of normalizations. In fact, (1) follows from a different attribution of the "ground energy", or minimal value, here interpreted as the intersection of the alternatives in each scenario, i.e., $\emptyset$ for the set of alternatives $\{\pi_{0,a}, \pi_{0,b}\}$ and $\{\mathtt{Y}\}$ for the set $\{\pi_{1,a}, \pi_{1,b}\}$. These two different choices represent unrelated normalizations, which open the way to incompatibility of preferences (opposite orders) between the two scenarios. We can better specify this observation by linking the decision contexts that generate the poset in Fig. 2a, i.e., the maximal elements of $\varrho_{\star}$, to Boolean algebras. Specifically, we map the Boolean algebra $\mathtt{B}^{(2)}$ to lotteries ℓ_u, for both $u \in \{0, 1\}$, and the Boolean algebra $\mathtt{B}^{(3)}$ to their combination $\ell_0 \vee \ell_1$:

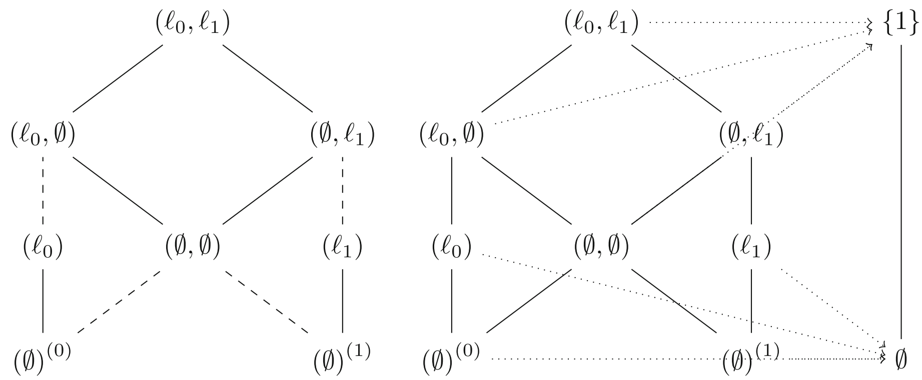
$$\begin{aligned}
\ell_0 &\mapsto \mathcal{B}_1 := (\wp\{\mathtt{R}, \mathtt{B}\}, \cup, \cap, \cdot \to \emptyset, \{\mathtt{R}, \mathtt{B}\}, \emptyset)\,, \\
\ell_1 &\mapsto \mathcal{B}_2 := \Big(\{\{\mathtt{Y}\}, \{\mathtt{R}, \mathtt{Y}\}, \{\mathtt{B}, \mathtt{Y}\}, \{\mathtt{R}, \mathtt{Y}, \mathtt{B}\}\}, \cup, \cap, \cdot \\
&\qquad \to \{\mathtt{Y}\}, \{\mathtt{R}, \mathtt{Y}, \mathtt{B}\}, \{\mathtt{Y}\} \Big), \\
\ell_0 \vee \ell_1 &\mapsto \mathcal{B}_{1,2} := (\wp\{\mathtt{R}, \mathtt{Y}, \mathtt{B}\}, \cup, \cap, \cdot \to \emptyset, \{\mathtt{R}, \mathtt{Y}, \mathtt{B}\}, \emptyset)\,,
\end{aligned} \tag{20}$$

where we have used the expression $\cdot \to \emptyset$ for the complement $\cdot^{\mathrm{C}}$. The ambiguity in Ellsberg's model entails the lack of a combination of the two Boolean algebras $\mathcal{B}_1$ and $\mathcal{B}_2$ to get $\mathcal{B}_{1,2}$. This combination would be feasible if we could distinguish the two algebras and associate them with substructures of $\mathcal{B}_{1,2}$. However, meta-reasoning (lower poset $\varrho_{\star}$ in Fig. 2a, ν-meta-states in Fig. 2b) does not allow for such a distinction. In this way, we get another instance of multiple implications (here, $\cdot \to \emptyset$ and $\cdot \to \{\mathtt{Y}\}$) already considered in Example 2, as well as the non-trivial effect of a "ground energy" labeling subsystems (here, $\emptyset$ and $\{\mathtt{Y}\}$ in the aforementioned implications) as in Angelelli (2017).

6.2 Wigner's model and data observability

While the labeling of contexts refers to lotteries in Ellsberg's model, in the Wigner's Friend scenario, it represents the two potential changes in the friend's state implied by the measurement, which are unknown to Wigner. As in the case of data-driven strategies, an agent is able to observe data (outcome of measurement in the first case, (big) data in the second one) that a super-observer (Wigner in the first case, a human agent in the second one) cannot.

Even for the Wigner's Friend scenario, before defining the formal correspondence with our framework (Sect. 6.2.2), we briefly discuss the source of uncertainty about features



(a) Using lower meta-states (b) Using ν-meta-states

Fig. 2 Representations of ambiguity in Ellsberg's three-color model using the two types of meta-states. Dashed lines (left figure) distinguish pairs in the poset $\varrho^\star$ that do not belong to the poset $\varrho_\star$. Dotted lines (right figure) depict the map ν linking $\varrho^\star$ to the Boolean algebra $V_{\exp} := \mathtt{B}^{(1)}$

extracted from non-explainable approaches using artificial agents.

6.2.1 Preliminary discussion

Let us consider an order relation on $\mathtt{f}_{(A)} := \{\mathtt{f}, \mathtt{f}_0, \mathtt{f}_1\}$ given by the inclusion of the indices. Specifically, we have $\mathtt{f} \preceq_{\text{data}} \mathtt{f}_0$ and $\mathtt{f} \preceq_{\text{data}} \mathtt{f}_1$. The label $\mathtt{f}$ refers to the word "factor" (or feature), namely a relevant attribute defining the decision context based on the observed data. The condition $\mathtt{f} \preceq_{\text{data}} \mathtt{f}_0$ means that $\mathtt{f}$ does not identify a decision context, while $\mathtt{f}_0$ does and, hence, is less ambiguous. Note the analogy with the urn model described in the previous subsection: $\mathtt{f}_0$ and $\mathtt{f}_1$ could represent two distinct decision contexts ℓ_0 and ℓ_1, resulting in two opposite orders.

The set $\mathtt{f}_{(A)}$ refers to the direct observation of data carried out by the artificial agent. So, we move to a second representation $\mathtt{F}_{(H)} := \left\{\mathtt{F}_{\{0\}}, \mathtt{F}_{\{1\}}, \mathtt{F}_{\{\emptyset\}}, \mathtt{F}_{\emptyset}\right\}$ to assess the knowledge possessed by the human agent about the AI's decisions. Specifically, the element $\mathtt{F}_{\{\emptyset\}}$ recognizes that the trained AI algorithm is in a defined but unknown decision context, and we consider the relations $\mathtt{F}_{\emptyset} \preceq_{\text{AI}} \mathtt{F}_{\{\emptyset\}} \preceq_{\text{AI}} \mathtt{F}_{\{u\}}$ for both $u \in \{0, 1\}$, where $\mathtt{F}_{\{\emptyset\}}$ is interpreted as "*the human agent knows that the AI knows the decision context*", while $\mathtt{F}_{\emptyset}$ is interpreted as "*the human agent knows that the AI does not know the decision context*". The subscript $\preceq_{\text{AI}}$ clarifies that the comparison refers to the AI's knowledge state.

As a consequence of the training with data, the AI updates its initial state to align with them. This update leads to the definition of a meta-state for the human agent, which reflects the changes in the AI's knowledge. Specifically, the knowledge of the AI's training prompts the human agent's knowledge to change from $\mathtt{F}_{\emptyset}$ to $\mathtt{F}_{\{\emptyset\}}$. This update acknowledges the alignment of the AI's outcomes with a feature in the (big) data, but the human agent remains unaware of the specific latent feature.

The relational structure defined by $\mathtt{F}_{(H)}$ can be encoded using the function ι introduced in (15) to provide an instance

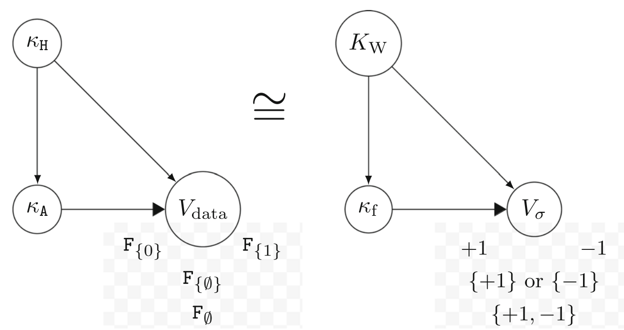


Fig. 3 Diagrammatic representation of a Wigner's friend scenario describing a high-level interaction of human and artificial agents in a data-driven scenario

of inequivalent knowledge representations. Specifically, we observe that

$$\iota(\emptyset) = \emptyset, \quad \iota(\{0\}) = \iota(\{1\}) = \{\emptyset\}, \quad \iota(\{0, 1\}) = \{\emptyset, \{0\}, \{1\}\}. \tag{21}$$

Then, the update from $\mathtt{f}$ to $\mathtt{f}_u$ for some $u \in \{0, 1\}$ prompts the update from $\mathtt{F}_{\iota(\emptyset)}$ to $\mathtt{F}_{\iota(\{u\})}$. We can describe the human agent's inability to explain the AI's outcome through the set difference $\Delta := \iota(\{0, 1\}) \setminus \iota(\{u\})$ as a means to represent the divergence between the full access to knowledge about the AI's decision contexts $\{0\} \vee \{1\} = \{0, 1\}$ (in $(\wp(\{0, 1\}, \subseteq))$ and the actual knowledge $\{u\}$. In this interpretation, from $\Delta \neq \emptyset$, we can say that the states $\mathtt{F}_U$ with $U \in \Delta$ are not accessible to the human agent.

This argument, which is graphically depicted in Fig. 3, is a basis for the specification of the formalism designed in this work for the Wigner's Friend scenario, as presented in the following subsection.

6.2.2 Representing uncertainty about data observability

For the dimension V_{data}, we choose a base set $\wp(\{\top\})$ to align with the original Wigner's Friend scenario. This encoding captures the effect of the spin measurement by Wigner's Friend as a transition from the two-dimensional space with basis $\{s_{+1}, s_{-1}\}$ of $\mathbb{C}^2$ to one of the one-dimensional spaces (with basis $\{s_{+1}\}$ or $\{s_{-1}\}$, respectively). Therefore, we define $V_{\mathrm{data}} := (\wp(\{\top\}), \subseteq)$ to represent the knowledge (observation or measurement) of a relevant feature (polarization) that allows value extraction from data (measured spin). Other definitions of V_{data} can be considered too, but it is worth noting that this choice also connects to the representation of Ellsberg's model in the previous subsection through the association of $\top$ with the knowledge of the ground energy, which distinguishes the inferences made in the two lotteries.

Data are observable for the AI (Wigner's Friend) but not for the human agent (Wigner); only the knowledge of the existence of an outcome observed by the AI (e.g., the conclusion of the training phase) is available to the human agent. Assuming that no other knowledge source besides data is needed to define the state of the AI, we set $\varrho_{\mathrm{A}} := V_{\mathrm{data}}$. While the acknowledgement of data value is given by $\pi_{\mathrm{data}}(\kappa_{\mathrm{H}})$, the acknowledgement of the value of the AI in κ_{H} is represented by the component $\pi_{\mathrm{AI}}(\kappa_{\mathrm{H}})$ along the dimension $V_{\mathrm{AI}} := (\wp(\varrho_{\mathrm{A}}), \subseteq)$. This abstracts the queries (measurements) $\wp(\kappa_{\mathrm{A}})$ that the human agent (Wigner) can ask the AI (Wigner's Friend) with state κ_{A}.

As a consequence of the actual observation of the outcome in the data-driven scenario, the states κ_{A} and κ_{H} are updated to encompass the existence of $\top$. The AI gains complete information about the relevant data dimension, leading to a change from $\pi_{\mathrm{data}}(\kappa_{\mathrm{A}}) = \emptyset$ to a new state κ'_{A} with $\pi_{\mathrm{data}}(\kappa'_{\mathrm{A}}) = \{\top\}$. We can express this update in accordance with Definition 5 by introducing $\psi_{\mathrm{A}} := (\{\top\}_{\mathrm{data}})$ and using the composition

$$\kappa_{\mathrm{A}} = (\emptyset_{\mathrm{data}}) \mapsto \kappa'_{\mathrm{A}} := \kappa_{\mathrm{A}} \vee \psi_{\mathrm{A}} = (\{\top\}_{\mathrm{data}})\,. \tag{22}$$

On the other hand, the realization of the training of the AI algorithm prompts a change in the knowledge state of the human agent; consistently with the transition $\kappa_{\mathrm{A}} \mapsto \kappa'_{\mathrm{A}}$, we describe

$$\kappa_{\mathrm{H}} = (\emptyset_{\mathrm{data}}, \wp(\emptyset)_{\mathrm{AI}}) \mapsto \kappa'_{\mathrm{H}} := (\emptyset_{\mathrm{data}}, \wp(\{\top\})_{\mathrm{AI}})\,, \tag{23}$$

which means that the human agent knows that the AI is aligned with the data provided for the training but is unable to directly query them. To formulate this limitation, we can express the transition from κ_{H} to κ'_{H} by introducing $\psi_{\mathrm{H}} := (\emptyset_{\mathrm{data}}, \{\{\top\}\}_{\mathrm{AI}})$ and looking at the explainability of the update $\kappa_{\mathrm{H}} \vee \psi_{\mathrm{H}}$ by ν-meta-states. As discussed in Sect. 6.2.1, we can use ι to generate ν-meta-states; an equivalent choice to analyze this scenario, where the update from κ_{H} to κ'_{H} does not affect the data dimension, is $\nu(\kappa_{\mathrm{H}}) := \iota \circ \pi_{\mathrm{AI}}(\kappa_{\mathrm{H}})$ where $\mathrm{AI} \in \underline{\kappa_{\mathrm{H}}}$, and $\emptyset$ otherwise. We find

$$\begin{aligned}(\kappa_{\mathrm{H}}, \iota(\kappa_{\mathrm{H}})) \vee (\psi_{\mathrm{H}}, \iota(\psi_{\mathrm{H}})) &= (\kappa_{\mathrm{H}} \vee \psi_{\mathrm{H}}, \iota(\kappa_{\mathrm{H}}) \vee \iota(\psi_{\mathrm{H}}))\\ &= ((\emptyset_{\mathrm{data}}, \wp(\{\top\})_{\mathrm{AI}}), \{\emptyset\})\\ &\neq ((\emptyset_{\mathrm{data}}, \wp(\{\top\})_{\mathrm{AI}}), \{\emptyset, \{\emptyset\}, \{\{\top\}\}\})\\ &= (\kappa_{\mathrm{H}} \vee \psi_{\mathrm{H}}, \iota(\kappa_{\mathrm{H}} \vee \psi_{\mathrm{H}})).\end{aligned} \tag{24}$$

So, the update leading to κ'_{H} is not explainable based on the previous definitions.

7 Discussion on implications for the assessment of data-driven strategies

As mentioned in the Introduction and Sect. 2, business performance, as an indicator of the value associated with a big data-driven strategy, can be seen as a consequence of actions implemented in line with the strategic value generated through big data analysis. This stage, which can be defined as *big data exploitation*, is itself a consequence of a preprocess aimed at interpreting and capitalizing information extracted from data, which is referred to as *big data capitalization* (Ylijoki and Porras 2019; Wu et al. 2022). Based on findings in the literature, the dimensional definition of value is limited to the final phase of big data exploitation, as its principal purpose is to specialize the different types of value associated with big data in the measurement of the generated value. This measurement is realized by comparing the observed and estimated value indicators, which provides a criterion to determine the success or failure of the implemented data-driven strategy.

The proposed framework acts as an additional layer that embeds value dimensions within a formal knowledge structure and, hence, encompasses specific types of uncertainty (Sect. 2.1) within the capitalization phase. In practice, these types of uncertainty undermine the data-driven strategy from the beginning due to the lack of proper assessments of useful information when data are in their raw state. This aspect is critical since the definition of data-driven strategies often takes place before raw data acquisition, when *a priori* evaluations of information or strategic value are not available (Manyika et al. 2011). On the contrary, the structures to represent knowledge, in particular inner states, could be implemented within the whole value transformation process to highlight potential inconsistencies, redundancies, or a lack of representative indicators. In this way, the present model could be adapted for strategy assessment over the different stages of data capitalization and exploitation. In addition, it could enhance the applicability of in-use big data frameworks by promoting their proper adoption and integration within specific phases of the value creation processes.

The analysis conducted by Ashton (2007) points out the need to understand and consider casual linkages in these indicators, as they may create redundancies and distortions in aggregate measures such as the *value creation index* (VCI). The VCI introduces new categories of information (e.g., innovation, alliances, and technology), which are combined and weighted along with firm performance (Ashton 2007). In this regard, the integration of inner state representations within the Big Data Value Chain is a means to formalize and, if needed, update the knowledge about the processes that lead to value creation. Specifically, the cyclical assessment of potential inconsistencies that require updating a knowledge representation may return evidence about the resource usage and skills needed at the different assessment stages, which is a premise for the study of (causal) relations among them.

The focus on the explainability of inner states' updates also matches the need to adapt the assessment methodologies from the capitalization to the exploitation phases. Indeed, value measurement in terms of business performance can benefit from both classical forecasting models and advanced analytic methods (Negro 2022), which allow for the estimation of probability laws to obtain informative statistics and indices. However, this focus may overlook the quantification of non-financial and organizational indicators. This requires the introduction of complementary methods to assess these value attributes. Our proposal fits into the design of such methods that, starting with knowledge representations, assess their compatibility and the need to update the dimensional structure. In line with current research streams that explore the measurement of non-classical forms of uncertainty in socio-economic and psychometric settings (Sect. 2.4), such methods formalize a type of configural invariance between different frameworks and studies (in the sense discussed in Sect. 3.2).

The conditions expressed in Sect. 4 are grounded in the theory of extended orders, which go beyond the classical Boolean structures underlying probabilistic models. In this way, we strengthen a common foundational basis for fuzzy logic (Della Stella and Guido 2012) and for the study of inequivalent representations of statistical systems (Angelelli 2017), as discussed in Sect. 5. In fact, the attention paid to element- and set-based representations is part of a broader investigation that explores reduction to or deviation from classical set membership through geometric models. Specifically, a geometric realization of the operational structure introduced in Sect. 4.1 arises from a limiting procedure for set functions, where a set-element correspondence derives from combinatorial (Angelelli and Konopelchenko 2018, Sect. 6–7) or algebraic constraints (Angelelli 2019). Other types of uncertainty can be explored from a geometric standpoint, including imprecise probabilities (Cuzzolin 2020, Sect. 2.2) and factor indeterminacy in multivariate methods (Rigdon et al. 2019, pp. 430–431). The latter shares a feature with the type of metrological uncertainty discussed in this work, namely, the multiplicity of models compatible with the same accessible or observed information. Indeed, although different in nature, both types of indeterminacy rely on families of model transformations that preserve a given structure. In (Rigdon et al. 2019), they generate different solutions consistent with the same covariance structure, while in our setting, we deal with order-preserving mappings with different operational structures (5).

In conclusion, this formalism lends itself to the generation of qualitative and quantitative criteria for meta-reasoning about knowledge, which may support reliable maturity assessments (Sect. 2.4). Further study should be devoted to the specification of our approach for the design of adaptive maturity models in line with the dynamics of capabilities and technological adoption.

8 Conclusion and future work

This work has laid the basis for a deeper investigation of knowledge uncertainty in data-driven strategies, which are becoming a dominant component of technological innovation with significant effects on socio-economic systems (Ndou et al. 2019). The contribution started by identifying different manifestations of uncertainty that may affect data-driven strategies, which have to be included in the intermediate and final evaluations of innovation initiatives for their proper analysis.

Here, we focused on the explainability of knowledge updates; future work will explore the structures to express and investigate the explainability of knowledge *states*. The relation between structural (logical and algebraic) and information-theoretic notions should be explored in more depth to exploit both qualitative and quantitative approaches for assessing uncertainty in epistemic representations. Specifically, normalization's role in representing information-theoretic inequivalence (Sect. 5.3) and the Ellsberg model (Sect. 6.1.2) can be translated into an information-geometric setting (Angelelli and Konopelchenko 2021, Sect. 5), prompting dedicated analyses within other entropy-based statistical models (Carpita and Ciavolino 2017) and fuzzy techniques (Ciavolino et al. 2014; Ciavolino and Calcagnì 2016).

In this way, we envisage practical advantages in the design of measurement tools to assess business maturity in the context of big data. Maturity has naturally been linked to business value through the assumption that an organization with a high level of maturity has a greater chance of turning potential value into created value. At the same time, maturity models and dimensional value models share many of the aspects that have been explored in this discussion. Future applications will define assessment questionnaires suited to the analysis of interactions between human agents and technologies for data-driven initiatives. Along with a chosen methodological architecture in terms of dimensions, these assessment tools should envisage the occurrence of multiple representations of the same latent construct with incompatible behaviors (e.g., different qualitative features of relations within the structural model).

A final aspect to consider, beyond the scope of this paper, is the critical analysis of the non-monotonic relation between data features and the value that can be generated. Having more data (higher volume) is not always synonymous with getting a higher value. From a perspective in which data are resources, we should look at the extent to which data, information, and knowledge representations could faithfully be represented as resources, or, instead, they require a multi-actor view. In this direction, Gervasi et al. (2023b) discussed how big data value chain models can be combined with new data governance models, such as the Data Mesh Dehghani (2022). The presented framework is likely to fit into this multi-actor scheme, where AI is an agent and is part of the tools that can be used by a Technology Mesh to combine data from different domains (Gervasi et al. 2023b). In this way, data-driven strategies should be considered on a different ground with respect to classical paradigms in software engineering. The variety of effects that could be generated by data-driven strategies and the use of AI tools, as we discussed, should be incorporated within management processes, as they need to support organizations in increasing their awareness of data-driven strategies and the reliability of generated value measurements.

Acknowledgements Mario Angelelli is a member of INdAM-GNSAGA (Istituto Nazionale di Alta Matematica - Gruppo Nazionale per le Strutture Algebriche, Geometriche e le loro Applicazioni).

Funding Open access funding provided by Università del Salento within the CRUI-CARE Agreement.

Data availability The nature of this article is purely theoretical and methodological; therefore, no data was generated, used, or referenced in the manuscript.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Abramsky S, Brandenburger A (2014) An operational interpretation of negative probabilities and no-signaling models. In: Springer (ed) Horizons of the mind. A tribute to Prakash Panangaden. American Physical Society, pp 59–75. https://doi.org/10.1007/978-3-319-06880-0_3

Abramsky S, Barbosa RS, Mansfield S (2017) Contextual fraction as a measure of contextuality. Phys Rev Lett 119(5):050504. https://doi.org/10.1103/PhysRevLett.119.050504

Ackoff RL (1989) From data to wisdom. J Appl Syst Anal 16:3–9

Aerts D, Geriente S, Moreira C, Sozzo S (2018) Testing ambiguity and Machina preferences within a quantum-theoretic framework for decision-making. J Math Econ 78:176–185. https://doi.org/10.1016/J.JMATECO.2017.12.002

Al-Sai Z, Husin H, Syed-Mohamad S, Abdullah R, Zitar R, Abualigah L, Gandomi A (2022) Big data maturity assessment models: a systematic literature review. Big Data Cognit Comput 7:2. https://doi.org/10.3390/bdcc7010002

Angelelli M (2017) Tropical limit and a micro-macro correspondence in statistical physics. J Phys A Math Theor 50(41):415202. https://doi.org/10.1088/1751-8121/AA863B

Angelelli M (2019) Complexity reduction for sign configurations through the KP II equation and its information-theoretic aspects. J Math Phys 10(1063/1):5086165

Angelelli M, Konopelchenko B (2018) Zeros and amoebas of partition functions. Rev Math Phys 30(09):1850015. https://doi.org/10.1142/s0129055x18500150

Angelelli M, Konopelchenko B (2021) Entropy driven transformations of statistical hypersurfaces. Rev Math Phys 33(02):2150001. https://doi.org/10.1142/s0129055x2150001x

Ashton RH (2007) Value-creation models for value-based management: review, analysis, and research directions. Adv Manag Account 16:1–62. https://doi.org/10.1016/S1474-7871(07)16001-9

Carpita M, Ciavolino E (2017) A generalized maximum entropy estimator to simple linear measurement error model with a composite indicator. Adv Data Anal Classif 11:139–158. https://doi.org/10.1007/s11634-016-0237-y

Cavanillas JM, Curry E, Wahlster W (2016) The big data value opportunity. In: New horizons for a data-driven economy: a roadmap for usage and exploitation of big data in Europe, pp 3–11. https://doi.org/10.1007/978-3-319-21569-3_1

Cervantes VH, Dzhafarov EN (2019) True contextuality in a psychophysical experiment. J Math Psychol 91:119–127. https://doi.org/10.1016/j.jmp.2019.04.006

Ciavolino E, Calcagnì A (2016) A generalized maximum entropy (GME) estimation approach to fuzzy regression model. Appl Soft Comput 38:51–63. https://doi.org/10.1016/j.asoc.2015.08.061

Ciavolino E, Salvatore S, Calcagnì A (2014) A fuzzy set theory based computational model to represent the quality of inter-rater agreement. Quality Quant 48:2225–2240

Corallo A, Crespino AM, Del Vecchio V, Gervasi M, Lazoi M, Marra M (2023) Evaluating maturity level of big data management and

analytics in industrial companies. Technol Forecast Soc Change 196:122826. https://doi.org/10.1016/j.techfore.2023.122826
Cuzzolin F (2020) The geometry of uncertainty: the geometry of imprecise probabilities. Springer, Cham. https://doi.org/10.1007/978-3-030-63153-6
Davey BA, Priestley HA (2002) Introduction to lattices and order. Cambridge University Press, Cambridge. https://doi.org/10.1017/cbo9780511809088
de Bruin T, Rosemann M, Freeze R, Kulkarni U (2005) Understanding the main phases of developing a maturity assessment model. In: ACIS 2005 proceedings—16th Australasian conference on information systems
Dehghani Z (2022) Data mesh: delivering data-driven value at scale, 1st edn. O'Reilly Media, Inc, Sebastopol
Della Stella ME, Guido C (2012) Extended-order algebras and fuzzy implicators. Soft Comput 16(11):1883–1892. https://doi.org/10.1007/s00500-012-0840-6
Doignon J-P, Falmagne J-C (2012) Knowledge spaces. Springer Verlag, Berlin Heidelberg
Dzhafarov EN, Kujala JV (2016) Context-content systems of random variables: the contextuality-by-default theory. J Math Psychol 74:11–33. https://doi.org/10.1016/j.jmp.2016.04.010
Elia G, Polimeno G, Solazzo G, Passiante G (2020) A multi-dimension framework for value creation through big data. Ind Mark Manag 90:617–632. https://doi.org/10.1016/j.indmarman.2020.03.015
Ellsberg D (1961) Risk, ambiguity, and the savage axioms. Q J Econ 75(4):643–669. https://doi.org/10.2307/1884324
Floridi L (2019) Establishing the rules for building trustworthy AI. Nat Mach Intell 1(6):261–262. https://doi.org/10.1038/s42256-019-0055-y
Fosso Wamba S, Akter S, Edwards A, Chopin G, Gnanzou D (2015) How"big data" can make big impact: findings from a systematic review and a longitudinal case study. Int J Prod Econ 165:234–246. https://doi.org/10.1016/j.ijpe.2014.12.031
Frauchiger D, Renner R (2018) Quantum theory cannot consistently describe the use of itself. Nat Commun 9(1):3711. https://doi.org/10.1038/s41467-018-05739-8
Geerts G, O'Leary D (2022) V-matrix: a wave theory of value creation for big data. Int J Account Inf Syst 47:100575. https://doi.org/10.1016/j.accinf.2022.100575
Gervasi M, Totaro NG, Fornaio A, Caivano D (2023a) Big data value graph: enhancing security and generating new value from big data. In: Buccafurri F, Ferrari E, Lax G (eds) Proceedings of the Italian conference on cyber security (ITASEC 2023), vol 3488. CEUR-WS, Bari. https://ceur-ws.org/Vol-3488/paper21.pdf
Gervasi M, Totaro NG, Specchia G, Latino ME (2023b) Unveiling the roots of big data project failure: a critical analysis of the distinguishing features and uncertainties in evaluating potential value. In: Proceedings of the 2nd Italian conference on big data and data science (ITADATA 2023), vol 3606, CEUR-WS, Naples. https://ceur-ws.org/Vol-3488/paper21.pdf
Gökalp MO, Gökalp E, Kayabay K, Koçyiğit A, Eren PE (2021) The development of the data science capability maturity model: a survey-based research. Online Inf Rev 46(3):547–567. https://doi.org/10.1108/oir-10-2020-0469
Greco S, Pereira RAM, Squillante M, Yager RR, Kacprzyk J (2010) Preferences and decisions. Springer, Berlin Heidelberg. https://doi.org/10.1007/978-3-642-15976-3
Gregor S, Martin M, Fernandez W, Stern S, Vitale M (2006) The transformational dimension in the realization of business value from information technology. J Strat Inf Syst 15(3):249–270. https://doi.org/10.1016/j.jsis.2006.04.001
Grover V, Chiang RHL, Liang T-P, Zhang D (2018) Creating strategic business value from big data analytics: a research framework. J Manag Inf Syst 35(2):388–423. https://doi.org/10.1080/07421222.2018.1451951
Gunning D, Stefik M, Choi J, Miller T, Stumpf S, Yang G-Z (2019) XAI-explainable artificial intelligence. Sci Robot 4(37):7120. https://doi.org/10.1126/scirobotics.aay7120
Günther WA, Rezazade Mehrizi MH, Huysman M, Feldberg F (2017) Debating big data: a literature review on realizing value from big data. J Strateg Inf Syst 26(3):191–209. https://doi.org/10.1016/j.jsis.2017.07.003
Halper F, Krishnan K (2013) TDWI big data maturity model guide: interpreting your assessment score. Technical report, TDWI Benchmark Guide
Halpern JY (2017) Reasoning about uncertainty. MIT Press, Cambridge. https://doi.org/10.7551/mitpress/10951.001.0001
Harding J (1996) Decompositions in quantum logic. Trans Am Math Soc 348(5):1839–1862. https://doi.org/10.1090/s0002-9947-96-01548-6
Henseler J, Ringle CM, Sarstedt M (2015) A new criterion for assessing discriminant validity in variance-based structural equation modeling. J Acad Mark Sci 43:115–135. https://doi.org/10.1007/s11747-014-0403-8
Henseler J, Ringle CM, Sarstedt M (2016) Testing measurement invariance of composites using partial least squares. Int Mark Rev 33(3):405–431. https://doi.org/10.1108/imr-09-2014-0304
Hussien AA (2020) How many old and new big data V's characteristics, processing technology, and applications (BD1). Int J Appl Innov Eng Manag 9:15–27
Ingusci E, Angelelli M, Sternativo GA, Catalano AA, Carlo ED, Cortese CG, Demerouti E, Ciavolino E (2023) A higher-order life crafting scale validation using PLS-CCA: the Italian version. Behaviormetrika. https://doi.org/10.1007/s41237-023-00209-y
Ishwarappa Anuradha J (2015) A brief introduction on big data 5Vs characteristics and Hadoop Technology. Procedia computer science, vol 48, pp 319–324. International conference on computer, communication and convergence (ICCC 2015)
Jamison DT, Lau LJ (1973) Semiorders and the theory of choice. Econometrica J Econom Soc 41(5):901–912. https://doi.org/10.2307/1913813
Lamba HS, Dubey SK (2015) Analysis of requirements for big data adoption to maximize IT business value. In: 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions). IEEE, pp 1–6
Laney D (2001) 3-D data management: controlling data volume, velocity, and variety. Technical report, META Group Res. Note
Maçada AC, Beltrame M, Dolci P, Becker J (2012) IT business value model for information intensive organizations. BAR Braz Adm Rev 9:44–65. https://doi.org/10.1590/S1807-76922012000100004
Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, Byers AH (2011) Big data: the next frontier for innovation, competition, and productivity. Technical report, McKinsey Global Institute
Mettler T, Rohner P, Winter R (2010) Towards a classification of maturity models in information systems. Management of the interconnected world. Physica-Verlag HD, Heidelberg, pp 333–340. https://doi.org/10.1007/978-3-7908-2404-9_39
Montequín V, Cousillas S, Ortega-Fernández F, Balsera J (2014) Analysis of the success factors and failure causes in Information & Communication Technology (ICT) projects in Spain. Procedia Technol 16:992–999. https://doi.org/10.1016/j.protcy.2014.10.053
Ndou V, Kalemi E, Elezaj O, Ciavolino E (2019) Toward a framework to unlock innovation from big data. Entrepreneurship, innovation and inequality. Routledge, London, pp 111–131. https://doi.org/10.4324/9780429292583-8
Negro L (2022) Sample distribution theory using Coarea formula. Commun Stat Theory Methods. https://doi.org/10.1080/03610926.2022.2116284

Nurgalieva N, del Rio L (2018) Inadequacy of modal logic in quantum settings. In: Proceedings of the 15th international conference on quantum physics and logic—QPL 2018, vol 287. EPTCS, ETH Zurich, pp 267–297. https://doi.org/10.4204/EPTCS.287
Patgiri R, Ahmed A (2016) Big data: the V's of the game changer paradigm. In: 2016 IEEE 18th international conference on high performance computing and communications. IEEE, pp 17–24. https://doi.org/10.1109/HPCC-SmartCity-DSS.2016.0014
Reggio G, Astesiano E (2020) Big-data/analytics projects failure: a literature review. In: IEEE, pp 246–255. https://doi.org/10.1109/SEAA51224.2020.00050
Rigdon EE, Becker J-M, Sarstedt M (2019) Factor indeterminacy as metrological uncertainty: implications for advancing psychological measurement. Multivar Behav Res 54(3):429–443. https://doi.org/10.1080/00273171.2018.1535420
Rindova V, Courtney H (2020) To shape or adapt: knowledge problems, epistemologies, and strategic postures under Knightian uncertainty. Acad Manag Rev 45(4):787–807. https://doi.org/10.5465/amr.2018.0291
Sozzo S (2017) Effectiveness of the quantum-mechanical formalism in cognitive modeling. Soft Comput 21(6):1455–1465. https://doi.org/10.1007/s00500-015-1834-y
Sozzo S (2020) Explaining versus describing human decisions: Hilbert space structures in decision theory. Soft Comput 24(14):10219–10229. https://doi.org/10.1007/s00500-019-04140-x
Tentori K, Bonini N, Osherson D (2004) The conjunction fallacy: a misunderstanding about conjunction? Cogn Sci 28(3):467–477. https://doi.org/10.1016/j.cogsci.2004.01.001
Tversky A (1969) Intransitivity of preferences. Psychol Rev 76(1):31–48. https://doi.org/10.1037/h0026750
Uddin MF, Gupta N (2014) Seven V's of big data understanding big data to extract value. In: Proceedings of the 2014 zone 1 conference of the American Society for Engineering Education. IEEE, Bridgeport, pp 1–5. https://doi.org/10.1109/ASEEZone1.2014.6820689
van de Wetering R, Mikalef P, Krogstie J (2019) Strategic value creation through big data analytics capabilities: a configurational approach. In: 2019 IEEE 21st conference on business informatics (CBI). IEEE, New York City. https://doi.org/10.1109/cbi.2019.00037
Vesset D, Girard G, Feblowitz J, Versace M, Burghard C, O'Brien A, Olofson CW, Schubmehl D, McDonough B, Woodward A, Bond S (2015) IDC MaturityScape: big data and analytics 2.0. Technical report, IDC
Vitari C, Raguseo E (2020) Big data analytics business value and firm performance: linking with environmental context. Int J Prod Res 58(18):5456–5476. https://doi.org/10.1080/00207543.2019.1660822
Weill P, Broadbent M (1998) Leveraging the new infrastructure: how market leaders capitalize on it. Harvard Business School Press, Boston
Wigner EP (1995) Remarks on the mind-body question. In: Philosophical reflections and syntheses. Springer, Berlin Heidelberg, pp 247–260. https://doi.org/10.1007/978-3-642-78374-6_20
Wu X, Liang L, Chen S (2022) How big data alters value creation: through the lens of big data competency. Manag Decis 60(3):707–734. https://doi.org/10.1108/MD-09-2021-1199
Ylijoki O, Porras J (2016) Perspectives to definition of big data. A mapping study and discussion. J Innov Manag 4:69–91. https://doi.org/10.24840/2183-0606_004.001_0006
Ylijoki O, Porras J (2019) A recipe for big data value creation. Bus Process Manag J 25(5):1085–1100. https://doi.org/10.1108/BPMJ-03-2018-0082
Zeleny M (1987) Management support systems: towards integrated knowledge management. Hum Syst Manag 7:59–70