Representations of epistemic uncertainty and its perception in data-driven initiatives

Mario Angelelli\textsuperscript{1,2}\textsuperscript{*} and Massimiliano Gervasi\textsuperscript{1}

\textsuperscript{1}CAMPI - Centre of Applied Mathematics and Physics for Industry, University of Salento, Lecce, 73100, Italy.
\textsuperscript{2}INdAM - Istituto Nazionale di Alta Matematica, GNSAGA, Italy.

Abstract

Emerging data-driven strategies, powered by the advent of AI, are reshaping decision-making processes, moving away from traditional reliance on direct data interaction. This paradigm shift introduces new challenges in assessing the impact of data-driven initiatives. To support these evolving methodologies, there is a crucial need for new models capable of describing the uncertainties stemming from limited data observability and the resulting ambiguities in decision-making. This contribution presents a novel conceptual model designed to deal with uncertainty in knowledge representations and reasoning about information transfer mediated by agents. Drawing from the multidimensional frameworks currently adopted to assess the value generated in data-driven initiatives, we provide an algebraic description of knowledge states and their dynamics. Specifically, we endow our model with a formal structure to compare and combine knowledge states; an update is represented through these combinations, and its explainability is based on their consistency in different dimensional representations. We discuss instances where inequivalent representations of knowledge can address some issues related to uncertainty about value dimensions. Furthermore, we can define a formal analogy with two scenarios that illustrate non-classical uncertainty in terms of ambiguity and reasoning about knowledge mediated by other (artificial) agents observing data.

Keywords: Knowledge representation, Uncertainty modelling, Ambiguity, Data-driven strategy, Value generation, Explainability
1 Introduction

The ongoing technological evolution enables the generation, acquisition, and storage of an ever-increasing amount of data. In parallel, the digital transformation is prompting interest in the adoption, in multiple contexts, of data-driven strategies supported by Artificial Intelligence (AI) to create strategic value in terms of reusable knowledge and decision support [1, 2]. We will refer to such processes as data-driven strategies. The growing interest is testified by the increasing scientific production [3, 4]. This interest seems to be driven by the “promise” of a high return on investment generated by the implementation of initiatives designed to exploit ad hoc techniques to analyse large masses 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 [5, 6]. In the literature, data, especially big data, are described in terms of Velocity, Variety, and Volume [7] using the original V’s of the Big Data (hereafter referred to as V’s). Since 2010, this definition of big data associated with a multi-dimensional characterisation has been enriched by new V’s. Currently, in the literature, it is possible to find more than fifty of them [8]. However, some of these features are not always intrinsic; they can be affected by a certain degree of subjectivity. Furthermore, they must be contextualised according to the data-driven strategy, which has occasionally led to confusing definitions [9], making it challenging to identify which V’s to accept and which to discard [10].

The focus on data stimulated by data-driven strategies should be analysed from the broader perspective of the Data-Information-Knowledge-Wisdom (DIKW) paradigm [5] and its evolving form, the Raw Data-Data Formats-Information-Knowledge-Wisdom (RDIKW) paradigm, as proposed by Wu et al. [11]. This is reflected in big data implementation projects in big data value chain models [6]. To support the identification of business value that can be generated through data-driven strategies, specific value-dimensional frameworks have been defined [12]. Specific technologies and skills are essential at various stages of the value chain to extract the intrinsic value of data [11]. Capabilities that can be 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 tools in the literature to assess organisations’ readiness for adopting and implementing big data initiatives [13–16].

The concept of extracting value from big data, often referred to as potential value, represents a probabilistic estimate of the expected value that an initiative could generate [9, 17, 18]. Actually, most data-driven initiatives fail to deliver the estimated value [19]. This discrepancy between expected and actual value suggests a need for a deeper investigation to better understand the relationship between the projected and observed value. In fact, only a few organisations have implemented a structured value measurement system to rigorously quantify the Return on Investment in their data-driven strategy [20].

The present work is motivated by the need to provide useful representations that can express some of these types of uncertainty. This need serves as a fundamental basis for integrating them into the execution of data-driven initiatives. A major source of uncertainty in this kind of initiative 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 become particularly evident in data analysis techniques based on Deep Learning (DL), as they are characterised by black box approaches that limit the interpretability of extracted information (e.g., selected features). Consequently, this undermines the value of the initiative at the knowledge and wisdom levels. Addressing this issue prompts the search for new approaches that enhance explainability in the adoption of tools relying on AI [21]. In this context, our goal is to provide a definition of explainability suitable for the type of uncertainty arising in data-driven initiatives, as this uncertainty can undermine the assessment or measurement of the generated value.

For this purpose, we adopt a structural approach for modelling and representing some of the aforementioned notions. We take advantage of formal analogies with other scenarios characterised by ambiguity, such as Ellsberg’s urn models [22], and explore inequivalent descriptions arising from different data observation levels in measurement settings, such as in Wigner’s friend experiment [23, 24]. Starting from a dimensional definition of knowledge value, we identify order-theoretic structures to evaluate the (lack of) explainability in terms of the update of a knowledge representation.

Our objective is to propose a high-level framework that serves both mathematical advancements and managerial implications. The proposed framework aims at supporting practical interpretations, rooted in basic order-theoretic notions that underlie most of the decision-making methodologies. We express this relational structure in algebraic terms to describe relevant logical properties. On the other hand, such a model should be endowed with enough structure to encompass inconsistencies and ambiguities that can arise within data-driven initiatives. In particular, we pay special attention to potential obstructions to the update of a knowledge representation in terms of explainability. This can support organisational processes where new evidence and intermediate evaluations during the initiative continuously lead to changing objectives and adopting an agile management methodology.

The abstraction level derived from this approach allows us to better identify conditions and structures that enable the description of inconsistencies or bounded resources in the update of a knowledge representation. Furthermore, it enhances the framework’s adaptability to different specifications across multiple contexts, ensuring scalability. This requirement is essential in practical applications to prevent ambiguities and inappropriate adoption of evaluation methodologies, which, as mentioned above, represent a factor that undermines a proper assessment of (big) data-driven initiatives [9].

The rest of this paper is organised as follows: in Section 2, we discuss the main motivations that drive our research of new methods to address epistemic uncertainty in data-driven initiatives, supported by evidence and analysis of case studies from the scientific literature. In Section 3, we provide a brief description of uncertainty scenarios that are relevant to our discussion, moving from multidimensional frameworks for data-driven strategies to ambiguity, measurements mediated by other agents, and deviations from classical or rational behaviours in decision-making. We present the central notions of the proposed formalism in Section 4, paying special attention to the representation of knowledge states as well as their updates and their explainability. The connections
between the issues identified in data-driven strategies and the proposed framework are explored in Section 5. Based on the description of such knowledge representations and their properties, we formalise the mapping between our formalism and the uncertainty scenarios in this Section 6. Conclusions and future work are discussed in Section 7.

2 Motivations

As anticipated in the Introduction, decision-makers are guided by the RDIKW paradigm [11] in the definition of data-driven strategies, but the evaluation of useful knowledge cannot be assessed a priori when data are in their raw state or, even more, have not been captured yet. This aspect is emphasised in big data initiatives, where the uncertainty of end results arises at the different levels of the RDIKW hierarchy. The peculiarity of (big) data-driven strategies is rooted in several factors, 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 [25].

As a potential consequence, the potential value is rarely matched by the actual value generated by completed data-driven initiatives in the literature, which can validate the potential estimates provided. On one hand, a study by Manyika et al. [26] predicted significant benefits, such as companies using big data in their product and service innovation processes saving up to 20–30% on product development costs and achieving 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 billion euros to 300 billion euros in new value [27]. On the other hand, despite this potential, there is a high failure rate for big data projects, reported to be as high as 85% by Reggio and Astesiano [19]. In a broader scope, Montequín et al. [28] conducted a study identifying and analysing 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 deficient 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. Such evidence suggests that the causal factors that can explain the failure of data-driven initiatives are not contextual and isolated but systemic [25].

This argument raises questions about the management of data-driven strategies. In the literature, it is uncommon to find unified models that consider data, human and artificial agents, capabilities, maturity, technologies, and generated value. Thus, understanding their interaction and developing a framework that manages these multiple factors is desirable. Frequently, agnostic and overly general definitions, such as those referring to the characteristics of big data, are compared with maturity models that are too specialised for specific contexts. However, due to opposite reasons, both models are generally not suitable for real cases and do not provide actual support to organisations during the decision-making process with interpretable information and insights.
At this point, it is legitimate to question whether the value observed through the choice of assessment tools can be identified with the value that is actually generated within an initiative. The adoption of different measurement settings reflects different perspectives and could provide different measures of value created. In turn, an initiative could not have a well-defined and unique value attribution, being both a failure and a success at the same time; this makes the notion of value a characteristic that is not intrinsic in the process or the initiative, and this specifies the notion of the subjectivity of the generated value mentioned in the Introduction. Such problems are known in earlier models analysed in the literature; in particular, the analysis conducted by Ashton [17] highlights the need to understand and consider casual linkages in the measures, as they may create redundancies and distortions in aggregate measures such as the value index.

Value measurement often focuses on firm performance, which can benefit from classical probabilistic forecasting models and advanced analytic methods [29] allowing for the estimation of probability laws to obtain informative statistics and indices. However, this focus often overlooks the quantification of non-financial and organisational indicators. This requires the introduction of “non-classical” or alternative indices to assess these value attributes. For instance, the Value Creation Index introduces new categories of information (e.g., innovation, alliances, technology) that are combined and weighted along with firm performance [17].

Another critical factor to consider in assessing and making decisions regarding data-driven initiatives is time. The value generated by data-driven strategies may take time to manifest a meaningful impact, and occasionally a big data initiative may stall due to a lack of foresight in understanding its potential [25]. Similar challenges arise when measuring maturity, and these issues can have a cascading effect on the value generated, especially in low-maturity organisations, where data-driven initiatives often aim to improve the overall maturity of the business to transition into a data-driven organisation.

Finally, a key point that characterises data-driven strategies is the role of AI. The limited interpretability of intermediate outcomes during the course of such processes also makes the adaptation of decision strategies during the initiative difficult. This is now opening up new questions regarding Explainable AI (XAI, see, e.g., Gunning et al. [21]), trustworthy AI [30], as well as human-centred and general-purpose AI, along with the notion of causality (see, e.g., [31, Ch. 4]). The lack of full control over data processing, whether in extracting information or during the training phase of automated tools (e.g., classifiers based on artificial neural networks), can reinforce assumptions underlying the data acquisition process or analysis methodology. This uncertainty has practical consequences for assessing the value captured from data, introducing ambiguity and the potential to reinforce biases in decision-making, as discussed by Popović et al. [32]. The lack of appropriate methods of measuring the value generated also arises when it is not possible to identify all possible positive or negative impacts caused by the data-driven initiative.

All these observations are supported by evidence from the literature regarding the analysis of case studies, which points out the needs and requirements for new methodologies and notions that can support the identification of uncertainty sources.
in value assessments. As a first step in this direction, the following sections introduce a formalism focused on the level beyond information in the RDIKW hierarchy, namely, knowledge. Building upon the multidimensional structure that underlies several value frameworks, supporting decision-making and management of data-driven initiatives, we explore basic structures to express comparison and update of knowledge states, as well as the explainability of such updates.

3 Value uncertainty scenarios

Now we introduce two well-known thought experiments in the domains of decision science and physics, respectively, that can support us in the representation of specific types of uncertainty in data-driven initiatives.

The first scenario, Ellsberg’s three-colour urn model, is a prime example of deviation from the Expected Utility Theory for preferences [33]. The second scenario, the Wigner’s Friend experiment, is drawn from quantum physics and focuses on the implications of non-classical (quantum) measurements. Even though the second scenario may seem out of the scope of this work, it can provide a useful analogy to select appropriate criteria to express (the lack of) explainability arising from the accessibility that different agents have to information and knowledge. We remark that the use of quantum formalism has proved useful in modelling the ambiguity resulting from Ellsberg scenarios [34] and, more generally, to represent deviations from classical behaviour in cognition [35], decision-making [36, 37], logic and operational theories [38, 39], and social sciences [40, 41].

Before discussing these scenarios, we clarify the connections with the concepts introduced in the previous section by analysing the current methodologies for dimensional representations of value generation from data.

3.1 Dimensional value representation in data-driven initiatives

Data have economic value, and the ability to extract value from data is becoming a key characteristic for organisations that need to remain competitive [42]. In order to define whether the implementation of a data-driven strategy has achieved its goals, it is necessary to measure the business value generated as a result. In the modelling requirements of Chen et al. [43], req-3: Integrate value discovery with value realization, or among the critical success factors of [44, 45], there is evidence of the need to monitor the value generated during big data initiatives, not only at the final stage, in the ultimate value measure of Gantz and Reinsel [46], but also throughout the value creation process, starting with the potential value that is budgeted to be generated.

The literature reveals that a dimensional structure is often used to characterise value, enabling its observation and measurement. For instance, Grover et al. [20] 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 (signalling effect or herding effect). Günther et al. [47] conducted a literature review on how organisations realise value from big data through “paths to value”. Fosso Wamba et al. [48] analysed the five criteria (dimensions) discussed by Manyika et al. [26] and
interpreted them as a different type of generated value. Elia et al. [12] 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 from the four dimensions of value in [49] and considered 22 types of information technology benefit. We point out a first generalisation of the value description, highlighting a hierarchical structure that acknowledges value through subsequent refinements: Elia et al. [12] defined dimensions from Value directions, which are in turn defined by key features.

Gregor et al. [49] conducted a large-scale survey involving more than a thousand organisations. The collected data also include information regarding ICTs in the organisation, 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. [49] are based on subjective assessments by the authors, which they acknowledge as a limitation of this approach. Secondly, 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 organisational and contextual changes, and vice versa. Although the framework presented by Elia et al. [12] 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., organisational) 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.

We can point out some instances of knowledge updating from the literature that correspond to the update of the dimensional architecture representing value. Specifically, we can consider the shift from the model proposed by Gregor et al. [49] to the one proposed by Elia et al. [12]. Another example of updating the dimensional architecture representing value can be found in [50]. Starting from the model in [49] (4 supra-dimensions) and [51] (1 supra-dimension), Maçada et al. [50] identified a new model that confirms the four supra-dimensions in [49], but permutes the sub-dimensions. This aspect is worth considering because it highlights that dimensional definitions are subjective and representative of a particular view, not only in terms of granularity but also in terms of classification.

These aspects can be analysed both from a theoretical and an empirical perspective, highlighting different factors that may influence value perception. Furthermore, in [47, Sect. 3.1.2], the authors discussed the current debate regarding the relation between algorithmic and human-based intelligence. We pay attention to this topic starting from Subsection 4.3 by exploring the relation between human and artificial agents (AI). In particular, we will see that this relation may affect the update of a value frame under incompatibility with strategic choices made a priori by analysts or scientists.
3.2 Ambiguity: urn models

Urn models encompass a large class of measurement designs that realise different forms of uncertainty, including non-probabilistic ones. An essential example is the three-colour urn model introduced by Ellsberg [22, pp. 653-654], which we briefly describe before connecting it with our value framework.

Let us consider an urn containing 90 coloured 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 ball or a yellow is drawn from the urn.

In this case, we introduce the symbol $\preceq_d$ to denote the preference relation of a decision-maker between the above-mentioned alternatives.

Notably, findings have revealed preferences

$$\pi_{0,a} \preceq_d \pi_{0,b} \quad \text{and} \quad \pi_{1,b} \preceq_d \pi_{1,a} \quad (1)$$

which contradicts the standard Subjective Expected Utility Theory [33]. Specifically, (1) violates the Sure-Thing Principle, one of Savage’s postulates that can be included in the Expected Utility Theory as a form of monotony between preferences and their expected utility. We refer to [34] for more details on this topic and for a proposed framework that uses quantum structures to formalise this form of ambiguity in decision-making.

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

3.3 Lack of observability: Wigner’s Friend

Wigner’s Friend is a pivotal thought experiment in quantum physics, originally conceived by E. Wigner (see, e.g., [23]) to highlight the crucial role of observers in measurements within the quantum domain. In this brief overview, we summarise the key aspects of this thought experiment that are relevant to our scope, directing readers to Frauchiger and Renner [24] 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 state of the system, namely, inconsistency associated with measurements that can be experimentally tested \[52\].

### 3.4 Evidence of non-classical behaviour in cognition and self-assessment

The aforementioned scenarios are practical examples that highlight the deviation from classically expected behaviour. However, similar phenomena are frequently observed in realistic scenarios. In these cases, the typical approach is to model statements based on first-order logic that makes use of classical connectives (disjunction, conjunction, and negation), which allows for defining events and measuring them using classical approaches based on Kolmogorov probability axioms and Bayesian criteria for conditioning and updating probabilities. Measurement tools to assess cognitive constructs, such as questionnaires, allow the estimation of such probabilities from frequencies, as well as the study of empirical correlations (and contingency tables for categorical variables), up to factor analysis and structural equation modelling \[53\] to estimate relations between abstract constructs. In the context of this work, such tools are specified in terms of maturity.

Assessing an organisation’s maturity in terms of BDA capabilities and resources is a key point in defining strategies for adopting and implementing data-driven strategies \[54\]. Corallo et al. \[55\] provided an analysis of the main maturity models, according to the three groups of attributes proposed by Mettler et al. \[56\]. 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 organisations and analysed in the literature follow specific standards \[57, 58\], the interpretability of the results is complex, and maturity models are often only descriptive or comparative in nature but rarely prescriptive. As highlighted in \[55\], maturity models can have different designs in terms of the number of dimensions and scoring method. The measured maturity will then depend on these factors and could be more or less reliable depending on them. In addition, it must be considered that the responses collected from respondents are influenced by the sample selected and the biases associated with it.

Given the relevance of the design of measurement settings to evaluate the attributes mentioned above, the investigation of potential sources of uncertainty in such assessment tools, especially maturity models, is essential for conducting proper analysis and getting useful insights from the acquired and expected 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 \( p(A \land B) \leq \min\{p(A), p(B)\} \) for events \( A, B \) weighted by the probability \( p(\cdot) \). This type of (conjunction) fallacy has been observed in questionnaires [59] and web searches [36]. Even in this case, the explanation of such phenomena can benefit from the Hilbert space representation of quantum mechanics [37].

Other deviations from rational (Bayesian) behaviour regard contextuality, namely, the dependence of an observed property on the whole experimental setting, which includes other simultaneously measured properties. This characterising aspect of quantum phenomena [38, 39, 60] extends to psychological measurements [61]. Empirical demonstrations of contextuality in psychological assessments have been conducted [40, 41] based on the verification (or the violations) of conditions implied by a classical model, namely, Bell-type and CHSH inequalities.

4 Model proposal: operational aspects of knowledge representations

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. 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 \( \psi_1 \) and \( \psi_2 \), we associate a labelled transition between them and denote it as

\[
\psi_1 \xrightarrow{\tau} \psi_2. \tag{2}
\]

In this way, given a state \( \psi_1 \), the class of all the possible transitions \( \tau \) originating from \( \psi_1 \) defines the possible inferences that an agent can make starting from \( \psi_1 \).

Motivated by the discussion in Section 2, now we include in our model the occurrence of multiple dimensions that guide the decision process and its evaluation. The minimal structure that supports decision-making criteria is a (pre-)order relation. So we provide the following:

**Definition 1** Let \( V \) be a set of partially ordered sets (posets, see, e.g., [62]). We label each element in \( V \) through an index set \( \mathcal{I} \), so we can express

\[
V := \{(V_i, \preceq_i) : i \in \mathcal{I}\} \tag{3}
\]

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 categorical product of the latent dimensions \( V_i, i \in \mathcal{J} \), which is the well-known Cartesian product for the category of sets (\( \text{Set} \)):

\[
\mathcal{V}_\mathcal{J} := \prod_{j \in \mathcal{J}} V_j. \tag{4}
\]

The corresponding value frame is then defined as the disjoint union (coproduct of sets) of these products, taken over all non-empty subsets \( \mathcal{J} \) of \( \mathcal{I} \):

\[
\mathcal{V} := \bigsqcup_{\emptyset \subset \mathcal{J} \subseteq \mathcal{I}} \mathcal{V}_\mathcal{J}. \tag{5}
\]
4.1 Modelling the role of observers: inner states

This subsection addresses the relevance of observers’ choices and the visibility of information.

**Definition 2** An inner state $\kappa$ is an element of the value frame $\varrho$. With a slight abuse of notation, we can equivalently express an inner state $\kappa$ as an indicator function: for a given $J \subseteq I$, we have

$$\kappa : J \rightarrow \varrho_J, \quad j \mapsto \kappa(j) \in V_j.$$  \hfill (6)

We emphasise that the use of the disjoint union is relevant to encompass contextuality, as can be seen in the special case when $V_i = \wp(S_i)$ for some set $S_i$ associated with each $i \in I$: a choice (6) does not always come from a choice of individual elements, namely, there does not always exist a subset $\kappa_\star \subseteq \bigcup_{i \in I} S_i$ such that $\kappa(j) = \kappa_\star \cap S_j$, $j \in J$. A formal approach to assessing the consistency of maps relating subsets and elements of a set with a given algebraic structure is presented in [63] and will be used in the following sections to represent ambiguity due to the lack of explainability.

**Definition 3** For each $H \subseteq J \subseteq I$, $\pi_H$ denotes the projection of $\varrho_J$ on the factor $\prod_{h \in H} V_h$, and we explicitly write $\pi_{J,i}$ to refer to the canonical projection on $V_i$, for each $i \in J$. An inner state $\kappa$ can also be viewed as a partial function from $I$ to $\coprod_{i \in I} V_i$: we will denote the domain and the codomain of this partial function $\kappa$ as $\text{dom}(\kappa)$ and $\text{cod}(\kappa)$, respectively.

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

This observation can be read in relation to the operator $K_i$ associated with the Agent $i$ in modal logic; the combination of the notions that are used to model knowledge structures (in particular, Kripke structures) in relation to Wigner’s friend extended scenarios proposed by Frauchiger and Renner [24] is discussed in [64, 65].

In the next section, we exploit the previous definitions to address the update of value frames, exploring the use of resources to produce knowledge from reusable information.

4.2 Comparison and composition

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 formalise concept analysis and knowledge structures [66].
This order relation can be described in terms of the composition of elements within each dimension \( V_i \). A stronger assumption that can be considered is the existence of an associative operation \( \oplus_i \) for each latent dimension \( V_i \) that is \textit{idempotent}, i.e., \( x \oplus_i x = x \) for all \( x \in V_i \) and \( i \in I \). This operation defines an order relation as follows:

\[
\forall a, b \in V_i : \quad a \preceq_i b \iff a \oplus_i b = b.
\] (7)

Such an operation, which combines the partial order \( \preceq_i \) and the notion of composition, is essential for the subsequent development and analysis of knowledge representations in the context of data-driven initiatives. The meaning of the statement \( a \preceq_i b \) is the following: let us consider any two agents \( A, B \) with inner states \( \kappa_A, \kappa_B \) respectively, such that \( i \in \kappa_A \cap \kappa_B, a = \pi_{\kappa_A,i}(\kappa_A) \), and \( b = \pi_{\kappa_B,i}(\kappa_B) \). Then, all the knowledge value recognised by Agent \( A \) along the dimension \( V_i \) is also recognised by Agent \( B \). In our perspective, the order relation represents the possible inferences that can be drawn by an agent using its knowledge resources. Regarding the totality of the order relation, partial orders allow highlighting non-equivalent representations of the value based on the same information. To clarify this aspect, in Section 5, we show an instance of this phenomenon using the results in [63].

In order to establish an algebraic structure for comparing different inner states and composing them, we endow the value frame \( \varrho \) with an order relation \( \preceq \) as follows: for all \( \kappa_1, \kappa_2 \) with \( \kappa_1 = \kappa_2 \), we consider \( \kappa_1 \preceq \kappa_2 \) if

\[
\forall i \in \kappa_1 : \pi_{\kappa_1,i}(\kappa_1) \preceq_i \pi_{\kappa_2,i}(\kappa_2).
\] (8)

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 ordering by taking into account the domains of inner states:

\[
\kappa_1 \preceq \kappa_2 \iff \kappa_1 \subseteq \kappa_2 \quad \& \quad \kappa_1 \preceq \kappa_2|_{\kappa_1}
\] (9)

where \( \kappa_2|_{\kappa_1} \) denotes the restriction of \( \kappa_2 \) to the subset \( \kappa_1 \), the domain of \( \kappa_1 \). We denote the poset \((\varrho, \preceq)\) as \( \varrho^* \).

\textbf{Remark 2} The domain extension (9) establishes a connection between different dimensional contexts \( \varrho_J \). This generates the order compatibility:

\[
\kappa_1|_{\kappa_1 \cap \kappa_2} \preceq \kappa_2|_{\kappa_1 \cap \kappa_2} \Rightarrow \kappa_1|_{\kappa_1 \cap \kappa_2} \preceq \kappa_2
\] (10)

The order (9) is a \textit{value-based} view, as it does not differentiate between different domains due to (10). Other orders can be associated with the set \( \varrho \); in particular, we can compare two elements only if they have the same support:

\[
\kappa_1 \subseteq \kappa_2 \iff \kappa_1 = \kappa_2 \quad \& \quad \forall i \in \kappa_1 : \kappa_1(i) \preceq_i \kappa_2(i).
\] (11)

This second order is a \textit{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 \( \varrho_* := (\varrho, \subseteq) \). Clearly, \( \varrho^* \) is an extension of \( \varrho_* \) since each pair \( \kappa_1 \subseteq \kappa_2 \) corresponds to a pair \( \kappa_1 \preceq \kappa_2 \) in \( \varrho^* \).
The two posets coincide only under the unidimensionality hypothesis. In the multi-
dimensional case, \( g \) and \( g^* \) act as lower and upper posets, respectively. This terminology
adapt the lower and upper probabilities (or belief and plausibility, respectively) that are
used to model imprecise probability, e.g., in Dempster-Shafer theory [67, Sect. 2.3-2.4] (also
see Cuzzolin [68] for a geometric view of these notions).

4.3 Self-reference

The distinguishing role of the domain can be used as a basis to consider different
orders starting with the same class of posets. In turn, this allows for modelling the
partiality of the composition (7). 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 \( \kappa_{\text{H}} \), respectively \( \kappa_{\text{A}} \), of a human agent
\( \text{H} \), respectively, and an artificial agent \( \text{A} \), the composition \( \kappa_{\text{A}} \oplus \kappa_{\text{H}} \) may not be feasible
when the knowledge of \( \text{A}'s \) inner state \( \kappa_{\text{A}} \) is partially accessible to \( \text{H} \).

Based on the scenarios described in Section 3, the type of uncertainty we want to
describe primarily arises through meta-reasoning, which, in our context, involves the
composition or comparison involving knowledge states along with value dimensions.
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
by Frauchiger and Renner [24].

Then, we can use the two posets \( g \) and \( g^* \) defined in Subsection 4.2 to represent
this form of meta-reasoning.

Definition 4 We extend \( g^* \) by adding a new poset \((g, \sqsubseteq_{\varphi})\) as a new dimension and repeating
the construction in (8)-(9). The corresponding poset
\[
K := (g \sqcup g \sqcap (g \sqcap g), \preceq_{\varphi})
\] (12)

with the order \( \preceq_{\varphi} \) induced by the aforementioned construction is referred to as the poset of
meta-states.

Remark 3 In Remark 2, we consider \( g \) as a lower poset, namely, \( \subseteq \sqsubseteq \preceq \) as relations on \( g \). On
the other hand, we can also consider a different poset of meta-states where \( g^* \) plays the role
of the lower poset, namely, \( \sqsubseteq \sqsubseteq \preceq \sqsubseteq \) as relations on \( g \). In particular, we can use a strictly
monotone function \( h : g \rightarrow \mathbb{R} \) to obtain a linear extension of \( \preceq \) as follows
\[
\kappa_1 \sqsubseteq \sqsubseteq \kappa_2 \Leftrightarrow h(\kappa_1) \leq h(\kappa_2).
\] (13)

This representation is used in the following sections to provide an information-theoretic view
of the state, where the extension \( \sqsubseteq \sqsubseteq \preceq \sqsubseteq \) is derived from an information measure.

4.4 Explainability as compositional existence

Finally, we use the notions introduced above to provide a formal definition of
explainability in our context.
Definition 5 A diagonal meta-state is an element of $\mathbf{K}$ that can be expressed as $(\kappa, \kappa)$ for some $\kappa \in \mathcal{K}$. The update of an inner state $\kappa \in \mathcal{K}^*$ is the composition $\kappa \lor \psi$ with another state $\psi \in \mathcal{K}^*$, when it exists.

Such an update is said explainable when $(\kappa, \kappa) \lor (\psi, \psi)$ also exists in $\mathbf{K}$ and, hence, is diagonal.

We can extend this definition by considering a new poset $V_{exp}$, an order-preserving mapping $\nu : \mathcal{K}^* \rightarrow V_{exp}$, and a poset of meta-states

$$K^{(\nu)} \coloneqq (\mathcal{K} \sqcup V_{exp} \sqcup (\mathcal{K} \sqcap V_{exp}), \preceq_\nu) \quad (14)$$

where $\preceq_\nu$ in (14) is obtained through the construction in (8)-(9). The elements of the poset $K^{(\nu)}$ are referred to as $\nu$-meta-states. A $\nu$-meta-state is diagonal if it has the form $(\kappa, \nu(\kappa))$ for some $\kappa \in \mathcal{K}$. An explainable update of $\nu$-meta-states is a composition (supremum) of diagonal elements in $K^{(\nu)}$ 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 updates of knowledge frames, in agreement with the attention paid to (non-)reusable knowledge in data-driven initiatives. When a combination $K \lor \kappa_\ell$ 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, thus limiting the opportunity to reuse in different contexts the resource carried by another inner state. An obstruction to the existence of such a state is the presence of multiple, non-equivalent value representations that cannot be directly discerned by the human agent. This is the situation we want to explore in order to describe interactions between human and artificial agents in structural terms.

Remark 4 The previous definition stresses the role of explainability in relation to accessible knowledge through the projections $\pi_{\mathcal{K},i}$ (see Definition 3). According to Remark 2, the role of the domain of inner states in the definition of $\mathbf{K}$, e.g., through the lower probability $\mathcal{K}$ in (11), relates the existence of an explanation for a meta-state in line with Definition 3 to the existence of the same set of projections $\pi_{\mathcal{J}}$, $\mathcal{J} \subseteq \mathcal{K}$, for both $\kappa$ and $\kappa \lor \psi$.

In the following sections, we analyse the definitions provided above in relation to inconsistencies that cannot be resolved through an explainable update of a meta-state. Within a data-driven initiative, this allows assessing whether data and other agents that can observe them generate reusable knowledge (an explainable update of a meta-state) or not.
5 Knowledge representations in relation to data-driven initiatives

5.1 Disjoint union and contextual comparison

We start the discussion of the application of the proposed notion to value assessment in data-driven initiatives by recalling the dimensional framework provided by Gregor et al. [49] and mentioned in Subsection 3.1.

The definition of each dimension is also detailed in [69, Tab. 10] in relation to big data; as an example, here we report three selected dimensions

\[ V_1 = \{f_1, f_2, f_3, f_4\}, \quad V_2 = \{m_1, m_2, m_3\}, \quad V_3 = \{b_1, b_2, b_3\} \]  

(15)

where the interpretation of these labels is detailed as follows:

- **V₁: Informational benefit**
  1. \( f_1 \): “Enabling faster access to information”
  2. \( f_2 \): “Enabling easier access to information”
  3. \( f_3 \): “Improving information for strategic planning”
  4. \( f_4 \): “Improving information accuracy”

- **V₂: Transactional benefit**
  1. \( m_1 \): “Savings in supply chain management”
  2. \( m_2 \): “Reducing operating costs”
  3. \( m_3 \): “Reducing communication costs”

- **V₃: Transformational benefit**
  1. \( b_1 \): “An improved skill level for employees”
  2. \( b_2 \): “Expanding organisational capabilities”
  3. \( b_3 \): “Improving organisational structure/processes”

For each dimension \( V_u \) in (15), \( u \in \{1, 2, 3\} \), ordinal assessments for the associated items can be composed to provide an overall evaluation of the knowledge of the dimension value, which is an element of a corresponding poset. These evaluations may be dependent due to the occurrence of underlying factors that constitute other dimensions. For instance, the element “improving organisational structure/processes” can belong to both “Transformational benefit” and “Transactional benefit”, depending on the way this improvement is carried out. This hypothesis can be read in connection with the work of Gregor et al. [49], where factor analysis is used as a tool to investigate the significant constructs within the survey response data and 22 dimensions are grouped into the 4 benefits. In particular, “Establishing useful links with other organisations” has factor loadings of 0.40 along the Strategic dimension and 0.33 along the Transformational dimension. If one uses the magnitude of factor loadings as a criterion of membership to different dimensions, then the narrow difference between these loadings might legitimise the definition of a structure where the benefit “Establishing useful links with other organisations” belongs to both the Strategic and the Transformational dimensions.
The search for deviations from unidimensionality, namely the occurrence of a unique latent factor that defines associated items or elements, is suited to the use of products and disjoint unions in our model. The disjoint union is needed to distinguish the same item or element within different dimensions or contexts and, hence, disambiguate different contents or interpretations of the element in different contexts. In data-driven initiatives, such a distinction acquires more relevance due to the intrinsic ambiguity in the definition of (big) data-characterising features [9].

The labelling induced by the disjoint product 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 [61]. To remove the dependence of such context, we can represent each element of \( \kappa \in \mathcal{G} \) as a tuple of pairs \((i, \kappa(i))\) and consider the equivalence relation defined by the projection on the second coordinate, so that \((i, \kappa(i))\) is identified with \((j, \kappa(j))\) if and only if \(\kappa(i) = \kappa(j)\).

In the present framework, different representations of the dimensions can address ambiguity in the definition of value dimensions since we can focus on \( \mathcal{V} := \wp \left( \prod_{u=1}^{3} \mathcal{V}_u \right) \) and redefine \( \mathcal{V}_1, \mathcal{V}_2, \mathcal{V}_3 \) as elements of \( \mathcal{V} \). In this way, we combine the membership in the “Establishing useful links with other organisations” benefit with distinct interpretations associated with the different dimensions. On the other hand, this representation does not take into proper account the lack of knowledge regarding the factor dimensionality, as we can always consider a refinement of the factor structure or higher-order constructs [53]. Meta-reasoning allows deriving inconsistency in this choice of dimensionality through the inclusion of a representative of the state itself, e.g., a dimension \( \mathcal{V}_{\text{dim}} := \wp (\mathcal{N}_\kappa) \) that only addresses the number of potential dimensions of the knowledge state. This lets us encode uncertainty on the dimensionality (\( \# \pi_{\text{dim}}(\kappa) > 1 \)), as well as the lack of specification (\( \text{dim} \notin \kappa \)). Furthermore, non-consistent models arise when the number of \( \# \kappa \notin \kappa(\text{dim}) \).

For this reason, it is essential to interpret the products in (4) within this context. These products, along with the canonical projections from \( \mathcal{G} \) into the individual dimensions, and the construction outlined in Subsection 4.4 let us relate to the aforementioned forms of uncertainty in data-driven initiatives. In the following subsection, we analyse in more detail the conditions resulting in the definition of meta-states and their implications.

5.2 Modelling bounded accessibility to dimensions

5.2.1 Non-classicality from multiple potential implications

Example 1 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 [62, Sect. 5.12]. It is well known that the implication \( A \to \cdot \) defined by \( A \to Y := Y \cup (A)^c \) is the upper adjoint of the conjunction \( A \cap \cdot \), namely, they are monotonic functions satisfying the relation
\[
A \cap Y \subseteq Z \Rightarrow Y \subseteq A \to Z, \quad A, Y, Z \subseteq \mathcal{S}.
\] (16)
where \( C \) 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 generalising classical logic to Heyting algebras and extended-order algebras in the context of fuzzy operators (see, e.g., [70] and references therein).
However, this construction presumes the knowledge of the whole set \( \mathcal{S} \) in order to be able to evaluate the complement \( \mathcal{C} \). In particular, we can model partial knowledge on \( \mathcal{S} \) by considering a class \( \{ \mathcal{S}_u, u \in \{1, \ldots, n\} \} \) of \( n \) potential spaces that define as many implications \( \rightarrow^u, u \in \{1, \ldots, n\} \).

**Remark 5** 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 of knowledge states \( \mathcal{K} \). When \( \mathcal{K} \) 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 characterise quantum logics (in particular, by replacing distributivity with modularity; see, e.g., [71]) and their extensions. In both alternatives, we get a non-classical behaviour inferred from bounded knowledge resources.

### 5.2.2 Inequivalent representations from accessibility boundary

For a statistical system, non-equivalent descriptions are often a hint of its non-trivial characteristics; see, e.g., ensemble non-equivalence [72] and temperature non-equivalence [63]. We investigate this phenomenon in our setting through the following example.

**Example 2** Let us take a class of dimensions \( \mathcal{V} \) and a distinguished dimension \( \mathcal{V}_i \) where the order relation \( \preceq_i \) is partial. Furthermore, suppose that the supremum \( a \lor_i b \) exists for all \( a, b \in \mathcal{V}_i \), which is an instance of (7). Another way to encode the relation \( \preceq_i \) is to introduce

\[
x \in \mathcal{V}_i \mapsto \iota_i(x) := \{ y \prec_i x : y \in \mathcal{V}_i \}.
\]

(17)

where \( a \prec_i b \) means \( a \preceq_i b \) and \( a \neq b \). This is a representation of value not based on a single valuation \( x \in \mathcal{V}_i \) but on the inferences that can be drawn from \( x \).

The statement \( a \preceq_i b \) implies that \( \iota_i(a) \subseteq \iota_i(b) \), so \( \iota \) is a strictly monotone mapping; in this sense, the order relation between compatible elements is preserved passing from \( \mathcal{V}_i \) to \( \mathcal{V}_i^\ast \) (i.e., an order morphism). However, these two representations differ when the composition structure (7) is considered: we can consider \( \lor \) (the supremum operation) and \( \cup \) as two operations satisfying (7) for \( \mathcal{V}_i \) and \( \mathcal{V}_i^\ast \), respectively. According to [63, Prop. 5.1], we know that \( (\mathcal{V}_i, \lor) \) is homomorphic to \( (\mathcal{V}_i^\ast, \cup) \) only if \( (\mathcal{V}_i, \preceq_i) \) is totally ordered. Since the order relation attributed to \( \mathcal{V}_i \) is not total, these different definitions of the dimension produce inequivalent results under composition.

### 5.2.3 Uncertainty on information measures

The lack of knowledge about the entire dimensional space (\( \mathcal{S} \) in Subsection 5.2.1, unique maximal elements in \( \iota(x) \) in Subsection 5.2.2) affects the quantification of the information content in knowledge representations. This lets us establish the analogy with the role of normalisation of subsystems discussed in [63]. Consider a state \( \kappa \) and the poset of non-negative reals \( (\mathbb{R}_{\geq 0}, \leq) \). Associate each dimension \( \mathcal{V}_i, i \in \kappa \), with a weight \( w_\kappa(i) \) for a given function \( w_\kappa : \kappa \to \mathbb{R}_{\geq 0} \). We can establish a connection with [63] by using a monotone function \( \nu : \mathcal{V} \to \mathbb{R}_{\geq 0} \), which was also employed for constructing \( \nu \)-meta-states in (14), to generate \( w_\kappa \). In particular, defining \( w_\kappa(i) := \)
\( \nu \left( \kappa_{\{i\}} \right) \), we take into account both the domain and the values of \( \kappa \); by restricting to the dimension \( g \), introduced in (11), we can associate each \( \kappa \in g \) with a weight \( w_{\kappa}(\kappa) \geq 0 \), so that \( \kappa_1 \subseteq \kappa_2 \) implies \( w_{\kappa}(\kappa_1) \leq w_{\kappa}(\kappa_2) \). 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 normalisation suited to our context, which only relies on the domain \( \kappa \).

Therefore, to assess the effect of the type of uncertainty we are investigating on information measures, we can take a knowledge representation where each \( V_i \) is isomorphic to the ordered poset \((\mathbb{R}_{\geq 0}, \leq)\), or, in accordance with the framework introduced in [63], a tropical semiring \((\mathbb{R}_{\geq 0}, \oplus = \max, 0)\). In this setup, each inner state \( \kappa \in g \), after normalisation, induces a probability distribution whose support is \( \kappa \). We consider as an information measure \( h: g \rightarrow \mathbb{R}_{\geq 0} \) the normalised Shannon entropy \([73, \text{Sect. 2.2}]\)

\[
h(\kappa) := \frac{H(\kappa)}{H_{\text{max}}(\kappa)} = -\frac{1}{\ln(\#(\kappa))} \sum_{t \in \kappa} \frac{w_t}{\sum_{u \in \kappa} w_u} \cdot \ln \left( \frac{w_u}{\sum_{t \in \kappa} w_t} \right)
\]

where \( \#(\kappa) \) is the cardinality of \( \kappa \). This function is widely adopted to quantify the uncertainty (or, dually, the information and the complexity) within a given distribution. The inclusion of the normalisation \( \ln(\#(\kappa))^{-1} \) derived from the maximum entropy achievable for a distribution with support \( \kappa \) takes into account the potential probability assignments to available dimensions. Such a dependence relates to the distinction between the value-based and domain-based views in Remark 2 and makes the normalised entropy undefined. We can infer bounds for this function when partial information is available on the support \( \kappa \), e.g., lower and upper approximations.

We also note that these two views express different levels of order for probability distributions that can be used to assess updates of knowledge states and their explainability. This kind of requirement also arises in information theory, e.g., when one considers the Kullback-Leibler divergence \([73, \text{Sect. 2.3}]\) 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 \( \kappa_1 \) and \( \kappa_2 \), respectively, the Kullback-Leibler divergence \( D_{KL}(p(\kappa_2) || p(\kappa_1)) \) from \( p(\kappa_1) \) to \( p(\kappa_2) \) can be evaluated only if \( p(\kappa_1) = 0 \) whether \( p(\kappa_1) = 0 \). This assumption of absolute continuity formalises 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 \( \kappa \) with new dimensions; furthermore, this extension is evaluated differently in the two posets \( g \) and \( g' \), which allows using lower and upper posets as a means to discriminate explainable updates, as anticipated in Subsection 4.3.

6 Modelling in data-agent interactions

Finally, we connect the uncertainty scenarios described in Subsection 3 to the proposed framework.
6.1 Ambiguity and data-agent interactions

6.1.1 Preliminary discussion

Before defining the connection between Ellsberg’s three-colour urn model and the proposed formalism, which will be specified in Subsection 6.1.2, we identify some preliminary analogies to contextualise the decision-making problem in the scope of this work. The decision-maker is represented by a human agent $\text{H}$, who has access to information regarding the value of the “Red” dimension; specifically, $\text{H}$ can assess the risk associated with “Red”, e.g., knowing its impact and probability. On the contrary, the only information possessed by $\text{H}$ about the remaining two colours lies on the “Black-Yellow” information dimension, which acknowledges the existence of the two colours and their cumulative probability weight ($\frac{2}{3}$).

An artificial agent, also referred to as artificial intelligence (AI) whose inner state is denoted as $\kappa_A$, can get access to data that lets it acknowledge the value of the “Black-Yellow” information dimension to a greater extent with respect to $\text{H}$. In particular, there may be a latent factor in the data that lets $A$ distinguish two “Black” and “Yellow” information dimensions. The human agent knows that $A$ is able to recognise 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 $A$ allows the extraction of value from the “Black-Yellow” information dimension, but this cannot prompt a change of knowledge state for $\text{H}$ that is able to discern the value of the “Black” and the “Yellow” evaluations.

We provide a diagrammatic depiction of this phenomenon in Figure 1.
6.1.2 Ambiguity and explainability of knowledge updates

Now we provide a formal correspondence between Ellsberg’s three-colour model and the present framework. We introduce the notation $B(n)$ for the Boolean algebra $(\wp(\{1,\ldots,n\}),\cup,\cap,\cdot,\wp(\{1,\ldots,n\}),\emptyset)$ of the power set $\wp(\{1,\ldots,n\})$. Consider two dimensions $B_0$ and $B_1$ isomorphic to $B(1)$. 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. We obtain $g \cong \wp(\{\top(0), \top(1)\})$ as Boolean algebras (with the associated operations left implicit), where we use the labellings (0) and (1) to distinguish the two lotteries as a result of the disjoint union. Then we focus on the explainability of the composition $\{\top(0)\} \lor \{\top(1)\}$ and obtain

$$\{\top(0)\} \lor \{\top(1)\} = \{\top(0), \top(1)\} \quad \text{in } g^*.$$  \hspace{1cm} (19)

On the other hand, in $g$, this composition is not defined.

An analogous result is obtained using $\nu$-meta-states as in Definition 5. Here we consider $\nu: g \rightarrow B(1)$ with $\nu(\kappa) = \{1\}$ if $\kappa = (\{\top(0), \top(1)\})$ and $\nu(\kappa) = \emptyset$ otherwise. From (19) and $\nu(\{\top(0)\}) = \nu(\{\top(1)\}) = \emptyset \subset \{1\} = \nu(\{\top(0), \top(1)\})$, we find that the composition of the two lotteries is not explainable. In Figures 2(a)-2(b), we provide a diagrammatic representation of the previous argument, denoting $\ell_0 := \{\top(0)\}$ and $\ell_1 := \{\top(1)\}$ to stress the link to the two lotteries in Ellsberg’s model.

We stress that a different view on an analogous phenomenon has been given in [63, Sect. 7] in relation to the emergence of different orderings for quantities characterising physical subsystems, as a consequence of different choices of their normalisations. 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 $\{Y\}$ for the set $\{\pi_{1,a}, \pi_{1,b}\}$. These two different choices represent unrelated normalisations, which open the way to incompatibility of preferences (opposite orders) between the two scenarios. We can better specify the connection with the present work by attaching a Boolean algebra to each node in the poset in Figure 2(a): specifically, we associate $B(1)$ to $(\emptyset, \emptyset)$, the Boolean algebra $B(2)$ to $(\ell_u, \ell_u)$ for both $u \in \{0, 1\}$, and the Boolean algebra $B(3)$ to $(\ell_0 \lor \ell_1, \{\top(0), \top(1)\})$. The lotteries $\ell_u$, $u \in \{0, 1\}$, corresponding to $(\pi_{u,a}, \pi_{u,b})$ with “ground energies” $\emptyset$ and
respectively, are associated with the following Boolean algebras

\[
\ell_0 \mapsto B_1 := (\wp\{R, B\}, \cup, \cap, \cdot \mapsto \emptyset, \{R, B\}, \emptyset),
\]

\[
\ell_1 \mapsto B_2 := (\{(R, Y), \{B, Y\}\}, \cup, \cap, \cdot \mapsto \{Y\}, \{R, Y, B\}, \{Y\}),
\]

\[
\ell_0 \lor \ell_1 \mapsto B_{1,2} := (\wp\{R, Y, B\}, \cup, \cap, \cdot \mapsto \emptyset, \{R, Y, B\}, \emptyset)
\]

where we have used the expression \(\cdot \mapsto \emptyset\) for the complement \(\cdot\). The ambiguity represented by Ellsberg’s model and formalised as above entails the lack of a combination of the two Boolean algebras \(B_1\) and \(B_2\) to get \(B_{1,2}\). This combination would be feasible if we could distinguish the two algebras and associate them with sub-structures of \(B_{1,2}\). However, meta-reasoning (lower poset \(\kappa^\star\) in Figure 2(a), \(\nu\)-meta-states in Figure 2(b)) does not allow for such a distinction. In this way, we get another instance of multiple implications (here, \(\cdot \mapsto \emptyset\) and \(\cdot \mapsto \{Y\}\)) already considered in Example 1, as well as the non-trivial effect of a “ground value” labelling subsystems (here, \(\emptyset\) and \(\{Y\}\) in the aforementioned implications) as in [63].

6.2 Wigner effect and data observability

Moving from the urn model described in the previous subsection, we can now formalise the analogy with Wigner’s Friend scenarios. While the labelling of contexts refers to lotteries in Ellsberg’s model, here 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 initiatives, 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 this scenario, before defining the formal correspondence between Wigner’s Friend experiment and our framework (Subsection 6.2.2, we briefly discuss the source of uncertainty on features extracted from non-explainable approaches using artificial agents, which provides an intuitive analogy with the scenario introduced in Subsection 3.3.

6.2.1 Preliminary discussion

Let us consider an order relation on \(f(A) := \{f, f_0, f_1\}\) given by inclusion of the indices. Specifically, we have \(f \preceq_{\text{data}} f_0\) and \(f \preceq_{\text{data}} f_1\). The label \(f\) refers to the word “factor” (or feature), namely, a relevant attribute defining the decision context based on the observed data. The condition \(f \preceq_{\text{data}} f_0\) means that \(f\) does not identify a decision context, while \(f_0\) does and, hence, is less ambiguous. Note the analogy with the urn model described in the previous subsection: \(f_0\) and \(f_1\) could represent two distinct decision contexts \(\ell_0\) and \(\ell_1\), resulting in two opposite orders.

The set \(f(A)\) refers to the direct observation of data, which is carried out by the artificial agent. So we move to a second representation \(F_{[H]} := \{F_{\emptyset}, F_{\{U\}}, F_{\{U\}}, F_{\emptyset}\}\) to assess the knowledge possessed by the human agent about the AI’s decisions. Specifically, the element \(F_{\emptyset}\) recognises that the trained AI algorithm is in a defined but unknown decision context, and we consider the relations \(F_{\emptyset} \preceq_{\text{AI}} F_{\emptyset} \preceq_{\text{AI}} F_{\{U\}}\) for both \(u \in \{0, 1\}\) where \(F_{\emptyset}\) is interpreted as “the human agent knows that the AI knows
the decision context”, while $F_{\emptyset}$ is interpreted as “the human agent knows that the AI does not know the decision context”. The subscript $\preceq_{AI}$ 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 $F_{\emptyset}$ to $F_{\{\emptyset\}}$. 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 $F_{|H|$ can be encoded using the function $\iota$ introduced in (17) to provide an instance of inequivalent knowledge representations. Specifically, we observe that

\[
\iota(\emptyset) = \emptyset, \quad \iota(\{0\}) = \iota(\{1\}) = \emptyset, \quad \iota(\{0, 1\}) = \emptyset, \{0\}, \{1\}
\]

then, the update from $f$ to $f_u$ for some $u \in \{0, 1\}$ prompts the update from $F_{\iota(\emptyset)}$ to $F_{\iota(\{u\})}$. 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 $\{0\} \lor \{1\} = \{0, 1\}$ (in $\wp(\{0, 1\}, \subseteq)$) and the actual knowledge $\{u\}$. In this interpretation, from $\Delta \neq \emptyset$, we can say that the states $F_U$ with $U \in \Delta$ are not accessible to the human agent.

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

### 6.2.2 Representing uncertainty on data observability

For the dimension $V_{\text{data}}$, 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 $\{s_+\}$ or $\{s_-\}$, respectively). Therefore, we define $V_{\text{data}} := (\wp(\{\top\}), \subseteq)$ to represent the knowledge (observation or measurement) of a relevant feature (polarisation) that allows value extraction from data (measured spin). Other definitions of $V_{\text{data}}$ 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 $Y$, which distinguishes the inferences made in the two lotteries $(\pi_{0,a}, \pi_{0,b})$ and $(\pi_{1,a}, \pi_{1,b})$.

Data are observable for the AI (Wigner’s Friend) but not by 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 $\varrho_A := V_{\text{data}}$. While the acknowledgement of data value is given by $\pi_{\text{data}}(\kappa_H)$, the acknowledgement of the value of the AI in $\kappa_H$ is represented by the component $\pi_{\text{AI}}(\kappa_H)$ along the dimension $V_{\text{AI}} := (\wp(\kappa_{\text{AI}}), \subseteq)$, which abstracts the queries (measurements) that the human agent (Wigner) can ask the AI (Wigner’s Friend).

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

$$\kappa_A = (\emptyset_{\text{data}}) \mapsto \kappa'_A := \kappa_A \lor \psi_A = (\{\top\}_{\text{data}}).$$

On the other hand, the acknowledgement of the training of the AI algorithm prompts a change in the knowledge state of the human agent; consistently with the transition $\kappa_A \mapsto \kappa'_A$, we describe

$$\kappa_H = (\emptyset_{\text{data}}, \wp(\emptyset)) \mapsto \kappa'_H := (\emptyset_{\text{data}}, \wp(\{\top\}))_{\text{AI}}$$

which means that the human agent knows that the AI is aligned with (big) data provided for the training but is unable to directly query the data. To express this limitation, we consider a $\nu$-meta-state, choosing $\nu := \iota$ as defined in (17). Even in this case, we can express the transition from $\kappa_H$ to $\kappa'_H$ by introducing $\psi_H := (\{\top\})_{\text{AI}}$; however, looking at the explainability of this update, we find

$$\begin{align*}
(\kappa_H, \iota(\kappa_H)) \lor (\psi_H, \iota(\psi_H)) &= (\kappa_H \lor \psi_H, \iota(\kappa_H) \lor \iota(\psi_H)) \\
&= (\wp(\{\top\}), \emptyset) \\
&\neq (\wp(\{\top\}), \emptyset, \emptyset) \\
&= (\kappa_H \lor \psi_H, \iota(\kappa_H \lor \psi_H)).
\end{align*}$$

(24)

So the update leading to $\kappa'_H$ is not explainable based on the previous definitions.
7 Conclusion and future work

This work has laid the basis to stimulate a deeper investigation of knowledge uncertainty in data-driven initiatives, which are becoming a dominant approach in technological innovation with significant effects on socio-economic systems. The contribution started with the recognition of different manifestations of uncertainty in the characterisation, decision, and assessment criteria that affect data-driven strategies, which have to be included in the intermediate and final evaluations of innovation initiatives for their proper analysis.

Future research for both methodological and application advances is prompted by the present investigation. Regarding the former, the proposed formalism pays special attention to set-element correspondence in continuity with the study of inequivalent descriptions of physical systems [63] but is part of a broader investigation of the information content in algebraic structures that explores reduction to or deviation from classical set membership. In this regard, geometric models can also support the complexity analysis of combinatorial families (specifically, sign configurations) associated with set functions along with their reduction to elements due to the algebraic constraints of subspace parameterizations [74]. This suggests further studies of geometric models to represent uncertainty.

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 with the aim of providing both qualitative and quantitative approaches to assess uncertainty in epistemic representations.

These developments can complement statistical analyses based on methodologies to estimate, compare, and evaluate the performance of models involving latent constructs. In particular, Structural Equation Models (SEMs) are widely applied in different domains, including psychology, economics, epidemiology, biology, and management. SEMs have been applied to evaluate the effect of management on performance, also in combination with information-theoretic approaches [75]. Furthermore, this class of techniques for multivariate analysis is complemented by management frameworks to create value from big data [76]. The use of SEMs to estimate abstract concepts, such as value and knowledge, leads to a particular type of uncertainty stemming from the distinction between constructs (models, methods) and conceptual variables (here, knowledge value). Our framework addresses this source of uncertainty in terms of inequivalent knowledge representations and, hence, is set up as a type of metrological uncertainty [77]. The integration of the proposed structures with SEMs could be enhanced by their geometric representation mentioned above, which shares common features with SEMs and the study of uncertainty in soft metrology [78].

In this way, we envisage practical advantages in the design of measurement tools to assess business maturity in the context of big data. Currently, has naturally been linked to the business value that an organisation intends to create and is able to create through the implementation of a data-driven strategy. It is reasonable to assume that an organisation 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. In fact,
these models are often associated with assessment questionnaires, which need to take into account many factors such as the choice of respondents, the various biases that may exist in the responses, the chosen architecture in terms of dimensions, hierarchies and weights, and the way in which the various agents, human or artificial, interface.

The characteristics of data-driven strategies must lead to modifications of classical paradigms, from project management to new models of data governance, which may include a multi-actor paradigm. In this direction, one can see in [25] how big data value chain models can be combined with new data governance models such as the Data Mesh [79].

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 higher value; from a perspective in which data are resources, it is fair to have the ability to access a monotonic representation through a proper definition of resources. For this reason, new data governance models include multi-actor paradigms. In this direction, [25] discussed how big data value chain models can be combined with new data governance models, such as the Data Mesh [79]. 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 the Technology Mesh to combine data from different domains [25]. In this way, it is clear that data-driven strategies must lead to a modification of classical paradigms and cannot be just an evolution of classical deterministic software engineering models. The variety of effects that could be generated by data-driven strategies and the use of AI tools, as in the presented framework, should be incorporated and formalised in project management processes, as they need to support managers and organisations in increasing their awareness of data-driven initiatives and the reliability of the measures of value generated.

Declarations

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Competing Interests

Financial interests: On behalf of all authors, the corresponding author states that there is no financial interests.

References

[1] Vitari, C., Raguseo, E.: Big data analytics business value and firm performance: linking with environmental context. International Journal of Production Research 58(18), 5456–5476 (2020) https://doi.org/10.1080/00207543.2019.1660822

[2] Brinch, M.: Understanding the value of big data in supply chain management and its business processes: Towards a conceptual framework. International Journal of
[3] Singh, V., Banshal, S.K., Uddin, A.: Scientometric mapping of research on 'big data'. Scientometrics 105 (2015) https://doi.org/10.1007/s11192-015-1729-9

[4] Cappa, F., Oriani, R., Perulfo, E., McCarthy, I.: Big Data for Creating and Capturing Value in the Digitalized Environment: Unpacking the Effects of Volume, Variety, and Veracity on Firm Performance. Journal of Product Innovation Management 38(1), 49–67 (2021) https://doi.org/10.1111/jpim.12545

[5] Ackoff, R.L.: From data to wisdom. Journal of Applied System Analysis 16, 3–9 (1989)

[6] Ylijoki, O., Porras, J.: A recipe for big data value creation. Business Process Management Journal 25(5), 1085–1100 (2019) https://doi.org/10.1108/BPMJ-03-2018-0082

[7] Laney, D.: 3-D Data Management: Controlling Data Volume, Velocity, and Variety. Technical report, META Group Res. Note (2001)

[8] Hussien, A.A.: How Many Old and New Big Data V’s Characteristics, Processing Technology, And Applications (BD1). International Journal of Application or Innovation in Engineering & Management 9, 15–27 (2020)

[9] Ylijoki, O., Porras, J.: Perspectives to Definition of Big Data: A Mapping Study and Discussion. Journal of Innovation Management 4, 69–91 (2016) https://doi.org/10.24840/2183-0606_004.001_0006

[10] Patgiri, R., Ahmed, A.: Big data: The V’s of the game changer paradigm. In: 2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS), pp. 17–24 (2016). https://doi.org/10.1109/HPCC-SmartCity-DSS.2016.0014 . IEEE

[11] Wu, X., Liang, L., Chen, S.: How big data alters value creation: through the lens of big data competency. Management Decision 60(3), 707–734 (2022) https://doi.org/10.1108/MD-09-2021-1199

[12] Elia, G., Polimeno, G., Solazzo, G., Passiante, G.: A multi-dimension framework for value creation through big data. Industrial Marketing Management 90, 617–632 (2020) https://doi.org/10.1016/j.indmarman.2020.03.015

[13] Al-Sai, Z., Husin, H., Syed-Mohamad, S., Abdullah, R., Zitar, R., Abualigah, L., Gandomi, A.: Big data maturity assessment models: A systematic literature review. Big Data and Cognitive Computing 7, 2 (2022) https://doi.org/10.3390/
bdec7010002

[14] Sundararajan, A., Hernandez, A., Sarwat, A.: Adapting big data standards, maturity models to smart grid distributed generation: Critical review. IET Smart Grid 3 (2020) https://doi.org/10.1049/iet-stg.2019.0298

[15] Vesset, D., Girard, G., Feblowitz, J., Versace, M., Burghard, C., O’Brien, A., Olofson, C.W., Schubmehl, D., McDonough, B., Woodward, A., Bond, S.: Ide maturityscape: big data and analytics 2.0. (2015)

[16] Halper, F., Krishnan, K.: TDWI big data maturity model guide: interpreting your assessment score. Technical report, TDWI Benchmark Guide (2013)

[17] Ashton, R.H.: Value-creation models for value-based management: Review, analysis, and research directions. Advances in Management Accounting 16, 1–62 (2007) https://doi.org/10.1016/S1474-7871(07)16001-9

[18] Ishwarappa, Anuradha, J.: A brief introduction on big data 5vs characteristics and hadoop technology. Procedia Computer Science 48, 319–324 (2015). International Conference on Computer, Communication and Convergence (ICCC 2015)

[19] Reggio, G., Astesiano, E.: Big-data/analytics projects failure: A literature review, 246–255 (2020) https://doi.org/10.1109/SEAA51224.2020.00050

[20] Grover, V., Chiang, R.H.L., Liang, T.-P., Zhang, D.: Creating strategic business value from big data analytics: A research framework. Journal of Management Information Systems 35(2), 388–423 (2018) https://doi.org/10.1080/07421222.2018.1451951

[21] Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., Yang, G.-Z.: XAI—Explainable artificial intelligence. Science Robotics 4(37), 7120 (2019) https://doi.org/10.1126/scirobotics.aay7120

[22] Ellsberg, D.: Risk, Ambiguity, and the Savage Axioms. The Quarterly Journal of Economics 75(4), 643–669 (1961) https://doi.org/10.2307/1884324

[23] Wigner, E.P.: Remarks on the Mind-Body Question. In: Philosophical Reflections and Syntheses, pp. 247–260. Springer, Berlin Heidelberg (1995). https://doi.org/10.1007/978-3-642-78374-6_20

[24] Frauchiger, D., Renner, R.: Quantum theory cannot consistently describe the use of itself. Nature communications 9(1), 3711 (2018) https://doi.org/10.1038/s41467-018-05739-8

[25] Gervasi, M., Totaro, N.G., Specchia, G., Latino, M.E.: Unveiling the roots of big data project failure: a critical analysis of the distinguishing features and uncertainties in evaluating potential value. In: The 2nd Italian Conference on Big Data
and Data Science - ITADATA2023 (2023). Accepted for publication

[26] Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A.H.: Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute (2011)

[27] Cavanillas, J.M., Curry, E., Wahlster, W.: The big data value opportunity. New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe, 3–11 (2016) https://doi.org/10.1007/978-3-319-21569-3_1

[28] Montequín, V., Cousillas, S., Ortega-Fernández, F., Balsera, J.: Analysis of the Success Factors and Failure Causes in Information & Communication Technology (ICT) Projects in Spain. Procedia Technology 16, 992–999 (2014) https://doi.org/10.1016/j.protcy.2014.10.053

[29] Negro, L.: Sample distribution theory using coarea formula. Communications in Statistics - Theory and Methods 0(0), 1–26 (2022) https://doi.org/10.1080/03610926.2022.2116284

[30] Floridi, L.: Establishing the rules for building trustworthy AI. Nature Machine Intelligence 1(6), 261–262 (2019) https://doi.org/10.1038/s42256-019-0055-y

[31] Pearl, J.: Causality. Cambridge University Press, New York, NY (2009). https://doi.org/10.1017/cbo9780511803161

[32] Popović, A., Turk, T., Jaklič, J.: Conceptual model of business value of business intelligence systems. Management 15, 5–29 (2010)

[33] Von Neumann, J., Morgenstern, O.: Theory of Games and Economic Behavior. Princeton University Press, Princeton, NJ. (1953)

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

[35] Aerts, D.: Description of many separated physical entities without the paradoxes encountered in quantum mechanics. Foundations of Physics 12, 1131–1170 (1982) https://doi.org/10.1007/BF00729621

[36] Sozzo, S.: Effectiveness of the quantum-mechanical formalism in cognitive modeling. Soft Computing 21(6), 1455–1465 (2017) https://doi.org/10.1007/s00500-015-1834-y

[37] Sozzo, S.: Explaining versus describing human decisions: Hilbert space structures in decision theory. Soft Computing 24(14), 10219–10229 (2020) https://doi.org/10.1007/s00500-019-04140-x
[38] Abramsky, S., Brandenburger, A.: An operational interpretation of negative probabilities and no-signalling models. In: Springer (ed.) Horizons of the Mind. A Tribute to Prakash Panangaden, American Physical Society, pp. 59–75 (2014). https://doi.org/10.1007/978-3-319-06880-0_3

[39] Abramsky, S., Barbosa, R.S., Mansfield, S.: Contextual fraction as a measure of contextuality. Physical review letters 119(5), 050504 (2017) https://doi.org/10.1103/PhysRevLett.119.050504

[40] Cervantes, V.H., Dzhafarov, E.N.: True contextuality in a psychophysical experiment. Journal of Mathematical Psychology 91, 119–127 (2019) https://doi.org/10.1016/j.jmp.2019.04.006

[41] Dzhafarov, E.N., Zhang, R., Kujala, J.: Is there contextuality in behavioural and social systems? Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 374(2058), 20150099 (2016)

[42] Fast, V., Schnurr, D., Wohlfarth, M.: Data-driven competitive advantages in digital markets: An overview of data value and facilitating factors. Lecture Notes in Information Systems and Organisation 48 LNISO, 454 (2021) https://doi.org/10.1007/978-3-030-86800-0_31

[43] Chen, H.-M., Kazman, R., Garbajosa, J., Gonzalez, E.: Big Data Value Engineering for Business Model Innovation, vol. 2017-January, pp. 5921–5930 (2017)

[44] Gao, J., Koronios, A., Selle, S.: Towards a process view on critical success factors in big data analytics projects. In: Americas Conference on Information Systems (2015)

[45] Saltz, J.S.: The need for new processes, methodologies and tools to support big data teams and improve big data project effectiveness. In: 2015 IEEE International Conference on Big Data (Big Data), pp. 2066–2071 (2015). https://doi.org/10.1109/BigData.2015.7363988

[46] Gantz, J., Reinsel, D.: Extracting value from chaos. Technical Report 2011, IDC IView (2011)

[47] Günther, W.A., Rezazade Mehrizi, M.H., Huysman, M., Feldberg, F.: Debating big data: A literature review on realizing value from big data. Journal of Strategic Information Systems 26(3), 191–209 (2017) https://doi.org/0.1016/j.jsis.2017.07.003

[48] Fosso Wamba, S., Akter, S., Edwards, A., Chopin, G., Guanzou, D.: How “big data” can make big impact: Findings from a systematic review and a longitudinal case study. International Journal of Production Economics 165, 234–246 (2015) https://doi.org/10.1016/j.ijpe.2014.12.031
[49] Gregor, S., Martin, M., Fernandez, W., Stern, S., Vitale, M.: The transformational dimension in the realization of business value from information technology. Journal of Strategic Information Systems 15(3), 249–270 (2006) https://doi.org/10.1016/j.jsis.2006.04.001

[50] Maçada, A.C., Beltrame, M., Dolci, P., Becker, J.: It business value model for information intensive organizations. BAR - Brazilian Administration Review 9, 44–65 (2012) https://doi.org/10.1590/S1807-76922012000100004

[51] Weill, P., Broadbent, M.: Leveraging the new infrastructure: how market leaders capitalize on it. Boston: Harvard Business School Press (1998)

[52] Proietti, M., Pickston, A., Graffitti, F., Barrow, P., Kundys, D., Branciard, C., Ringbauer, M., Fedrizzi, A.: Experimental test of local observer independence. Science advances 5(9), 9832 (2019) https://doi.org/10.1126/sciadv.aaw9832

[53] Ciavolino, E., Carpita, M., Nitti, M.: High-order PLS path model with qualitative external information. Quality & Quantity 49, 1609–1620 (2015) https://doi.org/10.1007/s11135-014-0068-x

[54] Wetering (b), R., Mikalef, P., Krogstie, J.: Strategic value creation through big data analytics capabilities: A configurational approach. In: Becker, J., Novikov, D. (eds.) 2019 IEEE 21ST CONFERENCE ON BUSINESS INFORMATICS (CBI), VOL 1. Conference on Business Informatics, pp. 268–275 (2019). https://doi.org/10.1109/CBI.2019.00037

[55] Corallo, A., Crespino, A.M., Del Vecchio, V., Gervasi, M., Lazio, M., Marra, M.: Evaluating maturity level of big data management and analytics in industrial companies (2022). (under submission)

[56] Mettler, T., Rohner, P., Winter, R.: Towards a classification of maturity models in information systems. (2009)

[57] Gökalp, M., Gökalp, E., Kayabay, K., Koçyiğit, A., Eren, P.: The development of the data science capability maturity model: a survey-based research. Online Information Review ahead-of-print (2021) https://doi.org/10.1108/OIR-10-2020-0469

[58] Bruin, T., Freeze, R., Kulkarni, U., Rosemann, M.: Understanding the main phases of developing a maturity assessment model. Australasian Conference on Information Systems (2005)

[59] Tentori, K., Bonini, N., Osherson, D.: The conjunction fallacy: A misunderstanding about conjunction? Cognitive Science 28(3), 467–477 (2004) https://doi.org/10.1016/j.cogsci.2004.01.001

[60] Abramsky, S., Brandenburger, A.: The sheaf-theoretic structure of non-locality
and contextuality. New Journal of Physics 13(11), 113036 (2011) https://doi.org/10.1088/1367-2630/13/11/113036

[61] Dzhafarov, E.N., Kujala, J.V.: Context–content systems of random variables: The contextuality-by-default theory. Journal of Mathematical Psychology 74, 11–33 (2016)

[62] Davey, B.A., Priestley, H.A.: Introduction to Lattices and Order. Cambridge University Press, Cambridge, UK (2002). https://doi.org/10.1017/cbo9780511809088

[63] Angelelli, M.: Tropical limit and a micro-macro correspondence in statistical physics. Journal of Physics A: Mathematical and Theoretical 50(41), 415202 (2017) https://doi.org/10.1088/1751-8121/AA863B

[64] Nurgalieva, N., Río, L.: Inadequacy of modal logic in quantum settings. In: Proceedings of the 15th International Conference on Quantum Physics and Logic - QPL 2018, vol. 287, pp. 267–297. EPTCS, ETH Zurich (2018). https://doi.org/10.4204/EPTCS.287

[65] Vilasini, V., Nurgalieva, N., Río, L.: Multi-agent paradoxes beyond quantum theory. New Journal of Physics 21(11), 113028 (2019) https://doi.org/10.1088/1367-2630/ab4fc4

[66] Doignon, J.-P., Falmagne, J.-C.: Knowledge Spaces. Springer, Verlag Berlin Heidelberg (2012)

[67] Halpern, J.Y.: Reasoning About Uncertainty. MIT press, Cambridge, MA (2017). https://doi.org/10.7551/mitpress/10951.001.0001

[68] Cuzzolin, F.: The Geometry of Uncertainty: The Geometry of Imprecise Probabilities. Springer, Cham, CH (2020). https://doi.org/10.1007/978-3-030-63153-6

[69] Elia, G., Polimeni, G., Solazzo, G., Passante, G.: A multi-dimension framework for value creation through big data. Industrial Marketing Management 90, 617–632 (2019) https://doi.org/10.1016/j.indmarman.2019.08.004

[70] Della Stella, M.E., Guido, C.: Extended-order algebras and fuzzy implicators. Soft Computing 16(11), 1883–1892 (2012) https://doi.org/10.1007/s00500-012-0840-6

[71] Harding, J.: Decompositions in quantum logic. Transactions of the American Mathematical Society 348(5), 1839–1862 (1996) https://doi.org/10.1090/s0002-9947-96-01548-6

[72] Touchette, H.: Equivalence and Nonequivalence of Ensembles: Thermodynamic,
Macrostate, and Measure Levels. Journal of Statistical Physics 159(5), 987–1016 (2015) https://doi.org/10.1007/s10955-015-1212-2

[73] Cover, T.M., Thomas, J.A.: Elements of Information Theory. John Wiley & Sons, Hoboken, NJ (2006). https://doi.org/10.1002/047174882X

[74] Angelelli, M.: Complexity reduction for sign configurations through the kp ii equation and its information-theoretic aspects. Journal of Mathematical Physics 60(7) (2019) https://doi.org/10.1063/1.5086165

[75] Ciavolino, E., Dahlgaard, J.J.: Simultaneous equation model based on the generalised maximum entropy for studying the effect of management factors on enterprise performance. Journal of applied statistics 36(7), 801–815 (2009) https://doi.org/10.1080/02664760802510026

[76] Ndou, V., Kalemi, E., Elezaj, O., Ciavolino, E.: Toward a framework to unlock innovation from big data. In: Entrepreneurship, Innovation and Inequality, pp. 111–131. Routledge, London (2019). https://doi.org/10.4324/9780429292583-8

[77] Rigdon, E.E., Becker, J.-M., Sarstedt, M.: Factor indeterminacy as metrological uncertainty: Implications for advancing psychological measurement. Multivariate Behavioral Research 54(3), 429–443 (2019) https://doi.org/10.1080/00273171.2018.1535420

[78] Angelelli, M.: Identifiability and dependence structure in factor models through monomial assignments in determinantal decompositions. Available at SSRN 4531595 https://doi.org/10.2139/ssrn.4531595

[79] Dehghani, Z.: Data Mesh: Delivering Data-Driven Value at Scale. 1.ed. O’Reilly Media, Inc. (2022)