informatics

MDPI

Review

How Data Quality Impacts Decision Outcomes: Scoping Review, Methodological Development, and Empirical Demonstration

Eric A. Andersson

Department of Communication, Quality Management and Information Systems, Sundsvall Campus, Mid Sweden University, 85170 Sundsvall, Sweden; eric.andersson@miun.se

Abstract

Public sector value generation increasingly relies on data provided by information systems, yet the impact of data quality (DQ) on decision-making remains underexamined. While prior research has identified associations between DQ and decision outcomes, strong causal evidence remains scarce. This study addresses the issue in two phases: first, through a scoping review of experimental research on the impact of value-critical DQ dimensions—accuracy, completeness, consistency, and timeliness—and second, through methodological development. Searches across Scopus, Web of Science, IEEE, and PsycInfo, supplemented by citation searching and Google Scholar to avoid inclusion bias, identified 20 studies after screening and full-text review. Rather than excluding studies based on methodological quality, the review critically examined existing experimental designs and their findings. The literature revealed substantial weaknesses, including unclear distinctions between objective and subjective variables, weak manipulations, ambiguous operationalizations, and small sample sizes. Consequently, existing understanding of DQ impact remains limited and often inconclusive. To address these limitations, the article proposes a formal model for DQ experiments and presents a proof of concept for completeness using open data. The findings suggest that completeness influences value appraisal and subsequent choice behavior, particularly under higher uncertainty and larger value differences between alternatives.

Keywords: experiment; data quality; information quality; method; information system; literature review; public sector

1. Introduction

Governments are increasingly using information systems (IS) data through open data, big data, and data-driven initiatives to support transparency, accountability, and decision-making—all tasks that require reliable data to be fruitful [1,2].

Although data quality (DQ) is widely recognized as critical in IS, it remains underspecified in practice. Both policy and research often focus on interoperability and legacy system issues, implicitly assuming that government data is already accurate and only needs to be shared [1,3–5]. This creates a gap between DQ beliefs and control over it: 82% of data executives in state and organizations report that low quality is a barrier to their work, yet only 40% have a system to track their losses [6]. Focusing on data access rather than intrinsic DQ risks weakening the data-to-decision-to-value chain and eroding public trust [2,7]. For example, the recent overestimation of UK inflation figures caused financial consequences and eroded institutional trust through faulty decisions [8]. Poor DQ has

Check for updates

Academic Editor: Antony Bryant

Received: 25 May 2026
Revised: 12 August 2026
Accepted: 19 August 2026
Published: 21 August 2026

Copyright: © 2026 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

well-documented financial repercussions and remains a persistent challenge for IS research and practice [9,10].

Despite widespread recognition of its importance, research on the impact of DQ on decision-making remains limited. Existing studies are predominantly correlational and rely on surveys that measure perceived accuracy or self-reported decision quality (see, e.g., [11,12]). Many follow the fitness-for-use tradition, which emphasizes consumer perspectives but also relies on survey-based correlational evidence [13,14]. Correlational designs have limited causal evidence because they do not adequately control for confounding factors [15,16]. Previous works also displays poor alignment between judgments of quality and actual quality [17–19]. This makes it unclear whether the relationships reported in such surveys reflect causal effects and what they actually measure. These limitations constrain our understanding of how DQ influences decisions and outcomes. They also reflect broader challenges in IS research, including weak theory, gaps between research and practice, and limited attention to the cognitive view of data [20–22]. Thus, this study aims to evaluate the current state of knowledge on the causal relationship between DQ and decision-making, with particular attention to how DQ is conceptualized, operationalized, and empirically tested in the literature.

This study uses a two-phase structure. First, it reviews the state of the art. Second, it uses these insights to redefine how to assess the impact of DQ on decision-makers experimentally. The study then provides a proof of concept for this methodological approach by investigating completeness using available data. More specifically, the review examines how prior research has investigated the causal impact of DQ. To identify methodological limitations and inform improvements, it evaluates both methods and results. This is necessary because previous reviews of DQ impact have been limited in scope and have not sufficiently evaluated the evidence (e.g., [10]). The analysis focuses on four of the most common DQ dimensions: accuracy, completeness, consistency, and timeliness [23,24]. This focus allows for greater analytical depth and reflects the expectation that the literature on less common dimensions is sparse. Because information use and evaluation depend on cognitive processes, the study also considers both actual and perceived quality [22]. This distinction is important because decision-makers may respond differently when they suspect that data are flawed than when they unknowingly rely on poor data.

The main contributions of this work are its assessment of the evidence and methodological advancement. By strengthening the empirical basis of DQ research, the study contributes to current and future understanding of the boundary conditions for DQ within IS.

The paper is structured as follows. First, we establish the theoretical foundation for why experimental studies are preferable to correlational approaches by outlining their respective strengths and weaknesses. We then provide an overview of DQ and define the study framework. Next, we describe the scoping review method, including the search strategy, findings, and evaluation approach. As this work seeks to assess the literature in substantial detail, the methodological analysis is incorporated into the review evaluation rather than used as an exclusion criterion. We then present and discuss the results of the investigation. Following this, we formally define the dimensions using set notation to address issues of measurement clarity and methodological consistency. Finally, we provide a proof of concept for one of the dimensions, discuss its implications, and conclude with broader discussion and final remarks.

1.1. Theoretical Foundation

1.1.1. Experiments and Causal Validity

IS has generally struggled to develop strong theories [20]. This may be due to its strong focus on applied science, with researchers often borrowing theories from source

disciplines to address contemporary IS and organizational issues rather than conducting basic research [25]. Recent research has suggested that experimental work, particularly online experiments, can support IS scholars in addressing this challenge [26].

All trained researchers recognize that "correlation is not causation," and disregarding this distinction has consequences. The strongest evidence for a scientific theory is its ability to make accurate predictions about new data, particularly through novel implications [27,28]. Predictive power indicates that a theory captures regularities in the world without being so closely fitted to existing data that it begins to describe noise [27]. Correlational evidence can suggest causal effects, as illustrated by studies examining empirical DQ and decision efficiency or economic losses [11,29]. However, such studies cannot establish that changes in the outcome resulted from deliberate changes in the independent variable. They therefore risk attributing causality to relationships that may instead reflect other factors. Conducting more studies can reduce this uncertainty, but few studies have examined the causal effects of DQ on decision-making.

Experiments provide the strongest design for testing causal effects. An experiment is not simply a nominal descriptor of an investigation but a formal feature of study design. Experiments are used when a study can (a) manipulate the independent variable(s) of interest and (b) randomize the smallest experimental unit, often the subject themselves [30].

The statistical models for (uncontrolled) regression and correlation are identical, but their designs differ substantially in the conclusions they support. Independent variable manipulation is required to move from observing a state in the world, which may reflect a spurious correlation, to demonstrating a dependence relation. In essence, if an intervention causes a variable, and only this variable, to change, it becomes less likely that another process explains the observed results. This is further strengthened by designs that include a control condition, or counterfactual, to rule out the possibility that the dependent variable would have changed in the same direction regardless [16,30]. By statistical theory, randomization induces distributional matching across conditions, minimizing selection bias and group differences. It also probabilistically excludes most confounding conditions, such as unmeasured differences in intelligence or socioeconomic status [30,31].

While the above generally describes between-subjects designs (BS), experiments can also use within-subjects designs (WS), in which each individual participates in all conditions and serves as their own control. This is a valid and frequently used approach that accounts for individual variance, but it requires adequate counterbalancing to reduce carryover and order effects. It can also violate the assumption of equal variance between pairs of conditions [32]. Common methods include complete counterbalancing, Latin square designs, and block randomization [33]. In short, WS designs can improve statistical power because structurally controlling BS variance requires fewer participants. However, by construction, WS designs do not follow the principles that allow randomized control trials to estimate the sample average treatment effect [30].

To illustrate the methodological issue, consider the two cross-sectional surveys by [11,34]. Ref. [11] uses a PLS-SEM model to study DQ as a formative construct, while [34] uses hierarchical regressions on the mean of an index of items. Both use correlational designs to assess the impact of DQ, sometimes referred to as information quality, and therefore risk faulty conclusions. The first study [11] uses no methodological or statistical controls, leaving any number of alternative confounding variables that could influence the model. The second study [34] regresses DQ on managerial satisfaction and uses IT investment as a control variable, but it is unclear whether this is appropriate. IT investment could plausibly be influenced by both DQ and satisfaction. Controlling for it could therefore create a spurious relationship. This is known as a collider variable. Neither confounding nor collider variables can be adequately addressed without an experimental setup because

the theoretical model itself may be incorrect [16,30]. Moreover, both studies rely on respondents' judgments of the quality of data in their respective domains, which can be difficult to measure accurately [17–19].

Figure 1 illustrates how correlational studies can yield faulty conclusions even when respondents accurately report their constructs. Case I illustrates reverse causality, Case II confounding, and Case III a collider. In all three cases, modeling the wrong relationship relative to the actual relationship can still produce statistically significant results.

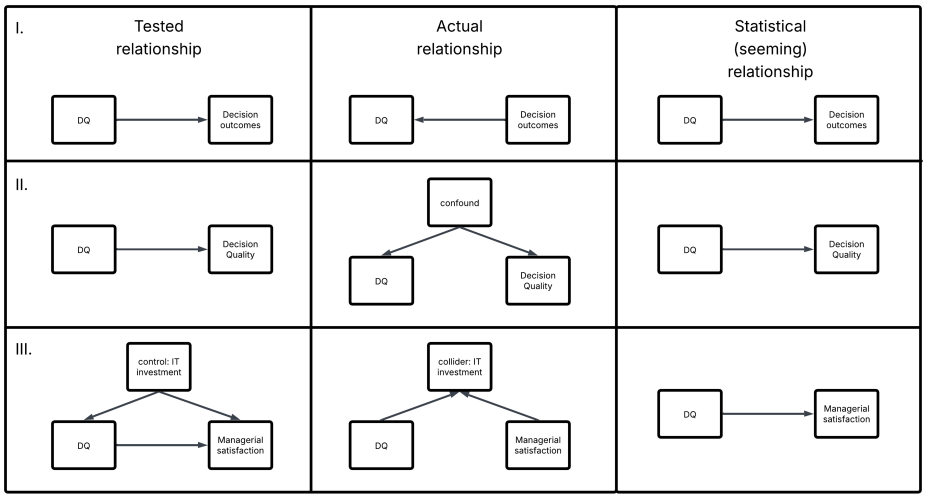


Figure 1. Examples of typical errors in correlational research.

Although large-scale organizational DQ studies often face design constraints, researchers studying the behavioral effects of DQ should directly manipulate DQ and randomly assign participants to conditions. When causal inference is the objective and an experimental design is feasible, it should generally be preferred because randomization provides stronger causal evidence than pseudo-experimental designs (manipulation without randomization) or non-experimental designs (without manipulation or randomization). However, these approaches remain valuable when experiments are infeasible. BS designs are preferable when adequate sample sizes are feasible, but WS designs remain appropriate when counterbalancing is properly implemented [33]. When manipulation or randomization is not possible, prospective designs are generally preferable to retrospective or cross-sectional approaches because they establish temporal order. Retrospective designs risk survivorship bias [35], while cross-sectional designs suffer from directional ambiguity as well as confounding and collider issues [16,30]. If surveys are used, researchers should address common method variance, for example, through single-factor tests, marker variables, or by incorporating both subjective and objective measures [36]. All variables previously shown or plausibly hypothesized to affect the model should be included [16,30]. Each experimental condition must also meet minimum sample size requirements. Although power estimates could be based on correlational evidence, they should also align with statistical theory. The central limit theorem suggests that samples of roughly 30 observations are sufficient for approximate normality and stable conclusions for many well-developed variables, with some sources suggesting that 20 observations may be acceptable [37–39].

1.1.2. Data Quality

In practice, DQ is deeply intertwined with organizational interpretation, performativity, and value construction. Individuals subject to data-driven evaluation may adapt their behavior to shape how data represent them, embedding performative dynamics

in the data themselves [40]. Organizational actors therefore do not passively consume data but actively participate in constructing what becomes accepted as organizational ground truth [41]. At the same time, data undergo continuous repair, framing, calibration, and amendment to make them valuable and actionable [42]. As shown by Mikalsen & Monteiro [43], organizations pragmatically judge when uncertain data are "good enough" for decision-making. Correlational surveys similarly risk conflating measurable DQ with subjective beliefs about organizational quality [11]. Addressing this distinction is therefore important for guiding practice.

There are numerous conceptualizations of DQ, with dimensions that sometimes overlap in their properties and sometimes share terminology despite differing in their conceptual foundations or operationalizations [10,44,45]. Most modern conceptualizations originate from the TDQM fitness-for-use tradition, also known as the information product perspective. In this tradition, information is treated as an organizational resource that can provide value to an end user if it is (a) free from deficiencies and (b) meets user needs [13,14]. DQ is divided into four categories: intrinsic, contextual, representational, and accessibility. Broadly, these categories represent properties of the data, properties of data in use, data structure, and within-system accessibility, respectively. Classically, and in many subsequent studies, dimensions within these categories are assessed through surveys of data consumers [11–13,34]. While useful and informative, this user-centered approach favors subjective measurement and the requirement that data meet user needs. Consequently, it gives less attention to intrinsic dimensions that can be assessed objectively, such as accuracy and completeness, and to the requirement that data be free from deficiencies.

Accordingly, this work takes an approach distinct from the fitness-for-use tradition: the theoretical, or data manufacturing, perspective [10,23]. This perspective focuses on internal aspects of DQ, or what other traditions refer to as intrinsic quality [13], as well as object-to-data semantic quality [46]. Its primary foundation is the work of Wand & Wang on ontology, which clarifies the relationship between real-world (RW) states and IS states [23]. They argue that errors occur when there is a mismatch between what data represent and the RW state they represent. Specifically, errors occur when 1. the IS does not represent an existing RW state (incomplete); 2. the IS state cannot be interpreted as a single RW state (ambiguous); 3. an IS state does not map to any RW state (meaningless); or 4. an RW state maps to a meaningless or incorrect IS state (garbling) Beyond these cases, the authors argue that errors can also arise from composition.

The present work focuses on four DQ dimensions: accuracy, completeness, consistency, and timeliness. This scope allows for an in-depth examination of some of the most common and relevant dimensions [10,47]. The following are some of the most prominent definitions of these dimensions.

The standard definition of accuracy is "The extent to which data are correct, reliable, and certified free of error." [13], p.31. It is also commonly referred to as correctness [24]. Common approaches to assessing accuracy include respondent opinion, gold standard comparison, source agreement, data element agreement, syntactic accuracy, and validity checks such as distribution comparison [24,48]. From an ontological perspective, inaccuracy occurs when either a meaningless state is treated as the RW state or the wrong IS state maps to the wrong RW state.

The standard definition of completeness is "extent to which data are of sufficient breadth, depth, and scope for the task at hand." [13], p.32. Common approaches to assessing completeness include respondent opinion, gold standard comparison, data source agreement, and the presence of data and variables [24,48]. From an ontological perspective, incompleteness occurs when an RW state is left unmapped in the IS, either as a missing value in a mapped instance or as a missing instance.

A typical definition of consistency is "the extent to which data satisfy constraints" [24], p.3. This most commonly means that the dataset must be internally consistent or in agreement with external standards [24,48]. From an ontological perspective, inconsistency occurs when two IS states that map to the same RW state are not identical, notwithstanding transformations.

The standard definition of timeliness is "The extent to which the age of the data is appropriate for the task at hand" [13], p.32, or "The extent to which data are promptly processed and up to date" [24], p.3. Common approaches to assessing timeliness include respondent opinion, time to availability, and log review (temporal examination of practice) [24,48]. Timeliness is closely related to data currency and can be understood as data being updated in time for use. From an ontological perspective, timeliness must therefore be understood temporally. A mapping may be correct at t = 0, but if the RW state changes at t = 1, the mapping may become invalid because it now maps to a meaningless state.

We use this framework to assess existing experimental research on the impact of DQ on decision-making.

2. Materials and Methods

The study used a scoping literature review to achieve its aim. In accordance with common review guidelines, we selected four databases for the searches: Scopus and Web of Science for their reputation as journal indexes, and IEEE Xplore and PsycInfo for their topical relevance [49–51]. These were supplemented by forward and backward citation tracking and snowball searches in Google Scholar to reduce inclusion bias [52,53]. The search keywords were (data quality OR information quality), (experiment OR RCT), decision*, and (accuracy OR correctness) OR completeness OR consistency OR timeliness. These were modified to meet the specific syntax requirements of each database.

Inclusion criteria: Studies were included if they experimentally examined the impact of DQ on decision-making using any of the specified DQ dimensions as search terms.

Exclusion criteria: Because the study aims to evaluate the field's methodologies alongside their results, no criteria were set for methodological adequacy, such as those discussed in Section 1.1.2. We excluded work that was not a peer-reviewed journal or conference article, was not available in English, or did not have an obtainable full text. Articles claiming an impact but not using an experimental design were also excluded. We limited the search to works published after 1990, a period characterized by substantial growth in IS research and the widespread adoption of personal computers, the internet, and LANs, which were key media for information transfer relevant to this discussion [25,54].

Search results: Initial searches on March 6th 2026 generated 296 results from Scopus, 135 from Web of Science, and 14 from PsycInfo. IEEE Xplore returned no valid results. Supplementary searches identified 52 articles for consideration. After abstract screening and duplicate removal, 47 unique articles were retained for further reading. Most exclusions at this stage involved studies using accuracy as a dependent variable to assess algorithms or algorithmic studies using the term experiment to describe designs that were more accurately classified as simulations or benchmark studies. Borderline cases that were unclear from the abstract were retained for full-text review. After full-text review, 20 articles met the inclusion criteria. The small final sample reflects the limited research in this area. The full-text inclusion and exclusion decisions are available at Zenodo (see data availability statement).

Evaluation method: The present study evaluated four aspects of study design to determine which findings were sufficiently reliable and valid for inclusion in the summative assessment of the impact of common DQ dimensions. Studies containing multiple experimental setups were evaluated separately for each configuration. First, each study

was classified as using a BS or WS design for the dimension under investigation. Although causality is generally less threatened in BS designs, WS designs are vulnerable to learning and order effects. Accordingly, we examined how stimuli were presented to identify these risks. For both BS and WS designs, full randomization was classified as adequate. In WS designs, random allocation among a limited set of stimulus orders was classified as borderline if ordering effects were also examined statistically. Similarly, studies using complete counterbalancing procedures, such as Latin square designs, were classified as adequate. At minimum, studies using non-randomized WS schemes were required to acknowledge and examine potential learning or order effects [33].

Second, we examined whether studies manipulated the intrinsic properties of the dataset itself (dataset manipulation; DM) or merely informed participants that the data were deficient without altering them (data quality information; DQI). DQI manipulates perceived quality while leaving objective quality unchanged, thereby targeting subjective beliefs rather than the informational properties themselves. This distinction affects the conclusions that can be drawn from these studies. Moreover, DQI designs are particularly vulnerable to demand characteristics, especially in WS settings where repeated exposure may make the manipulation salient to participants.

Third, we evaluated total sample size and the number of participants allocated to each condition. As stable statistical estimates generally require at least 30 observations per group [37,38], studies with 30 or more participants per condition were classified as adequate. Because some recent work suggests that 20 participants may be sufficient [39], studies with 20–29 participants per condition were classified as borderline, while smaller samples were classified as potentially unreliable.

Fourth, we assessed how each DQ dimension was operationalized. Specifically, we examined (a) whether the operationalization aligned with the dimension's definition and conceptual treatment presented in the previous section, indicating whether the intended construct was actually examined; and (b) the measurability of the operationalization based on its level of measurement. Although ordinal DQ manipulations may capture qualitatively meaningful differences, they restrict the statistical analyses available and provide less insight into the magnitude of manipulation effects. Across all criteria, unreported methodological features that prevented evaluation were treated as weaknesses in study design.

The classifications of previous research limitations are not intended to dismiss their contributions but to identify where their approaches limit causal inference and how they can be extended through more rigorous designs.

Following this evaluation, we synthesized findings from the most methodologically robust studies, namely those that best minimized causal ambiguity, while interpreting findings from weaker designs with caution based on their specific limitations and collecting them for methodological development [16,30,37]. The analytical criteria used in this assessment are summarized in Table 1 .

Table 1. Study aspects evaluation.

Judgment	Design (Variable, BS/WS)	Manipulation (DM/DQI)	Group Sizes	Definitional Adherence	Measurability
Adequate designs	Full randomization of conditions/trials, Latin square	Objective/subjective differentiation	>29	Proper mapping to errors [23]	Interval, ratio
Borderline designs	Partial randomization of trials to set orders	Objective/subjective acknowledgment	20–29	Ambiguous mapping	ordinal/qualitative
Weaker designs	Self-selection/No order effects examination	Objective/subjective neglect	<20	Wrong mapping (garbling)	nominal
	unreported	unreported	unreported	indeterminable	indeterminable

3. Results

3.1. Accuracy/Correctness

Much of the existing research on accuracy focuses on DQI rather than manipulating the underlying data themselves. These studies therefore primarily assess reactions to perceived differences in quality. This distinction is important because informing participants that data are of poor quality differs fundamentally from requiring them to work with objectively low-quality data. Treating these approaches as equivalent risks conflating the effects of objective and perceived quality in both analysis and reporting. A related weakness of many designs is the tendency to operationalize accuracy as reliability. Whereas accuracy refers to the extent to which data truthfully correspond to their referents [46], reliability reflects how frequently data have been accurate over time, that is, the probabilistic expectation [23].

Excluding studies with less robust sample sizes, the available evidence from DQI designs suggests that quality statements increase perceived quality and confidence in decisions based on that information. These effects are stronger when participants are also informed that the information source is trustworthy, which may further influence decision outcomes [55–57]. However, these findings do not directly demonstrate the effects of accuracy itself; rather, they reveal that individuals incorporate new informational cues into their evaluations and decisions. In two studies [55,57], participants never observed the underlying data and received only DQI. They instead judged how they would act based on information indicating that the unobserved data were poor. This design therefore captures responses to accurate DQI rather than to the underlying data, making it difficult to attribute the observed effects to DQ itself. Another study that provided participants with data found that experts were more capable than novices of integrating DQI into their judgments [58]. Although these designs demonstrate that people respond to DQI, they do not directly provide evidence that DQ affects their actions.

On the DM-design side, no studies employed adequate sample sizes, although the available findings tentatively suggest improvements in task and decision quality when accuracy is manipulated directly [59,60]. Two studies combined DQI with no-DQI performance, representing a pseudo-DM design. They found that decision quality improved when information was more accurate without participants being informed of its quality, while the presence of such information increased decision time [61,62].

However, these findings must be interpreted in light of definitional adherence. First, operationalizing accuracy as the probability of correctness does not fully capture the correspondence between a dataset and its referent. This reflects an older conceptualization of accuracy that does not distinguish it from related DQ dimensions in the way more recent frameworks do [55,57,61,62]. Second, ref. [59] operationalizes accuracy using a scaled error measure. The authors define this as $(x/(\text{correct value} - \text{most distant bottom value})$, $BV < x < CV$, then aggregate it across all x. This formulation has several problems. It constrains error to a fraction of the correct value and can produce negative accuracy. It can also produce accuracy values above one when BV and CV are close in magnitude, and it is undefined rather than equal to one when there is no variance. The measure also cannot accommodate negative values. Consequently, it behaves more like a fractional measure and is functionally closer to completeness, where quality decreases as fewer instances are correctly represented.

In summary, experimental evidence on the effects of accuracy on performance is limited and has weaknesses in sample size and adherence to modern definitions. Although the number of available studies is small, their findings generally align with the normative expectation that higher quality improves performance. The results of the accuracy review are presented in Table 2. For each study, the main results of the dimensional analyses are presented as statistically significant positive relations (+), statistically significant negative

relations (−), or non-significant results (n.s.) when investigated. We summarize our main judgments in the table, with more extensive analysis provided in the Zenodo material.

Table 2. Overview of previous experimental research designs for Accuracy.

Work	Design ($\bar{n}$)	Accuracy Operationalization (Descriptor)	Main Results from Increased Accuracy	Summary Judgment
[63]	BS, DQI (-)	0–100 (alt. above/below average) (reliable)	(1/8 sig.) top choice changed (4/8 sig.) ignoring DQI	unknown sample, wrong mapping
[58]	BS, DQI (59, 19 (uneven)) (35)	0–1 | provided DQI or not (reliable)	Novices ignore, experts integrate experts integrate	low sample, wrong mapping adequate
[56]	WS, DQI (37)	85, 70, 55% previous performance of advice	(+) decision confidence (sig.) reported value	wrong mapping
[64]	BS, DQI (16)	green, yellow, red (quality tag)	ignored IDQ having DQI (+) time to decide	low sample, ordinal/unmeasurable
[61]	BS, DM [1] (15) (15)	0–1, BS: presence of DQI ("percentage of accuracy")	(+) novice task accuracy (n.s.) time (novice) (n.s.) expert task accuracy (−) time (experts)	low sample, unknown definition, wrong mapping
[59]	BS, DM (10)	x/(correct value-most distant value)	(+) decision quality [3]	low sample, wrong mapping
[62]	BS, DM [1] (27)	0–1 | having DQI (chance of correct)	(+) [2] task accuracy (−) time to decide	borderline sample, wrong mapping
[60]	BS, DM (20)	0.6, 0.8, or 0.9 (chance to receive truth)	(+) worker performance (−) superior oversight	borderline sample
[57]	WS, DQI (25 | 27)	5-stage chance true (credibility)	(+) judged information accuracy unjudicable credibility—average accuracy	borderline sample, wrong mapping, demand characteristics risk
[55]	WS, DQI (37 | 88)	low, high chance true (credibility)	(+) judged quality (+) judged accuracy > (+) trustworthy (+) judged quality if source reliable	wrong mapping, ordinal, demand characteristics risk

[1] A special case, performance is judged against displayed quality, making no DQI condition perfect dataset quality. [2] DQ novices performed worse. [3] Supply chain decisions. BS = Between-subjects design. DM = Dataset Manipulation. DQI = Data quality information. WS = Within-subjects design.

3.2. Completeness

Contrasting with accuracy, most completeness assessments manipulate the dataset by providing participants with only a subset of the full material. However, it remains unclear whether completeness affects decision-making, as studies with adequate sample sizes failed to reject the null hypothesis. The operationalization closest to the definition of completeness did find an effect on decision quality but only allocated 10 participants to each condition, making the finding unreliable and in need of replication [59]. One study without broader methodological limitations found no effect on preferred gamble, but it is unclear whether this resulted from the smaller manipulation, which hid only one of six dice-roll payoffs rather than testing several levels of completeness [65]. Several studies also manipulated completeness qualitatively, for example, by adding or removing details. Although this is a reasonable way to create differences in completeness, it makes the magnitude of the manipulation difficult to measure and may result in only weak differences between conditions, potentially confounding the results [66–68]. The results of the review are presented in Table 3.

Table 3. Overview of previous experimental research designs for Completeness.

Work	Design ($\bar{n}$)	Completeness Operationalization (Descriptor)	Main Results from Improved Completeness	Summary Judgment
[67]	WS, DM (29)	lacking status, number, or direction of enemy aircraft	(n.s.) simulation perf.	borderline sample, qualitative
[59]	BS, DM (10)	present elements/all elements	(+) decision quality [1]	low sample
[68]	BS, DM (20)	green, yellow, red ± detailed breakdown (quality tag)	(+) threat judgment	borderline sample, qualitative
[65]	BS, DM (94)	5/6 or 6/6 dice outcomes known	(n.s.) preferred option (n.s.) trust in advisor (n.s.) switch advisor	adequate
[66]	BS, DM (43)	advice ±"strong argument"	(n.s.) investment judgment	qualitative

[1] Supply chain decisions.

3.3. Consistency

The literature contains few studies examining different forms of inconsistency. The findings are mixed: consistency does not appear to improve argument persuasiveness, but medium levels of information consistency may make individuals more predictable [55,69]. Two studies also produce conflicting results because they use different scoring rules for consistency, with one assuming a linear improvement and the other suggesting that medium consistency is more favorable [55,57]. Moreover, these studies derive consistency from DQI, including information about accuracy and source reliability, rather than from inconsistencies in the data themselves. As with accuracy, this makes it unclear whether the findings generalize to decisions based on objectively inconsistent data. The only other study, which most closely follows the definition of internal inconsistency [23], suggests that consistency does not affect decisions, but its small sample limits the conclusions that can be drawn [59]. The results of the review are presented in Table 4.

Table 4. Overview of previous experimental research designs for Consistency.

Work	Design ($\bar{n}$)	Consistency Operationalization (Descriptor)	Main Results from Improved Consistency	Summary Judgment
[59]	BS, DM (10)	conflicts/all cells for value & format	(n.s.) decision quality [1]	low sample
[69]	BS, DM (47 | 52)	expert contrast/support target	(n.s.) target argument persuasiveness	qualitative
[57]	WS, DQI (25 | 27)	source reliability—accuracy agreement	(+) test-retest consistency (+) consistency between analysts	borderline sample, ambiguous mapping, demand characteristics risk
[55]	WS, DQI (37) (88)	source reliability—accuracy agreement	(+, medium) test-retest consistency [2] (+, medium) test-retest consistency	ambiguous mapping, demand characteristics risk

[1] Supply chain decisions. [2] Redoes the analysis of [57] with their scoring rule (MANE) to achieve the same result as their own on test-retest consistency.

3.4. Timeliness

Despite being common in practice, experimental studies of timeliness appear to be limited to machine learning algorithms rather than human decision-makers [24,70]. For example, one study operationalizes dataset timeliness as the proportion of tuples that have not been replaced, such that Timeliness = (n − R)/n, where R is the number of replacement instances introduced for reclassification and n is the total dataset size [70]. This operationalization captures timeliness by varying the proportion of older instances replaced with newer ones and is broadly consistent with the present notion of garbling, in which a value fails to update and becomes outdated or meaningless. The measure therefore provides an overall indication of dataset currency and, if fitness-for-use is assumed to depend largely on currency, is consistent with timeliness as commonly conceived: data being available when needed. However, the operationalization may also overlap with accuracy because unreplaced tuples can become inaccurate when their underlying RW states change. Thus, it remains unknown whether timeliness affects human decision-making. Although timeliness affected algorithmic performance, it is unclear whether these effects would translate to decision-makers using the algorithm's outputs. Theoretically, timeliness should affect judgment when decisions are made using outdated information under changed conditions. Its effects are also likely to depend on context, as some tasks are more sensitive to change and some objects are less stable.

3.5. Experiments Using Non-Standard Dimensions

A few studies have examined data and information quality using non-standard terminology, focusing on holistic quality rather than individual dimensions. As a result, they provide limited clarity regarding which additional variables were manipulated and how they can be measured. These include partially significant results, with 3 of 5 tests indicating an effect of data-cleaning operations on task performance but not task completion time (BS, DM, 73) [71]. Another study found a significant effect of contextual DQ, contrasting relevant, complete, and aggregated data with their opposites, on task accuracy and completion time (BS, DM, 10) [72]. Improvements in information quality through increasingly relevant social interactions also improved the accuracy of personality judgments (BS, DM,

56) [73], while another study found similar effects only when interaction duration was sufficiently long (BS, DM, 78) [74]. Finally, objective, analytical, and persuasive negative product reviews significantly affected purchase risk, purchase intention, and product attitude compared with casual, brief, and subjective negative reviews (BS, DM, 64) [75]. Broadly, these manipulations are too broad to support detailed IS planning or identify specific mechanisms, but they nevertheless suggest that overall DQ and information quality can improve performance.

4. Discussion

4.1. Discussion of Review Findings

The literature on experimental DQ effects is sparse. Existing studies mostly highlight that DQI can influence judgment and confidence but provide limited evidence that DQ itself affects performance, while the mechanisms underlying these effects remain underexplored. Non-significant findings may also reflect weak manipulations rather than true null effects, as it is often unclear how much the qualitative and ordinal variables actually differed between conditions. Of 19 BS experiments across 16 studies, only 11 had adequate sample sizes. Among seven WS experiments across four studies, four had adequate sample sizes. Their controls included complete randomization of items (3), reversing question order for half the participants (2), and partial randomization into two versions (2). None fully counterbalanced carryover and order effects, and only one discussed learning effects, limiting exposure to three of five conditions [67].

Four studies also instructed participants to use DQI according to specified decision strategies, making their results measures of adherence to instructions rather than natural responses to DQ as they might occur in practice [58,61,63,64]. Only three studies examined management-adjacent tasks: supply chain decisions [59], oversight [60], and job attractiveness ranking [58]. Only one of these had an adequate sample size and found that experienced decision-makers, but not novices, tended to use rather than ignore DQI [60].

One finding suggests that the ontological model may require an extension: Ge & Helfert [59] incorporate error magnitude into their operationalization of accuracy. This captures an aspect overlooked by most operationalizations, as errors of the same size in data elements can have very different consequences. For example, an ice cream truck selling 500 or 50,000 units in a day would be inaccurate by one tuple in either case, but the magnitude of the error differs substantially. Error magnitude should therefore be incorporated into accuracy measurement.

Overall, evidence for DQ's impact is weak. Studies have limited causal strength, often fail to distinguish perceived from actual quality, use small samples, and employ operationalizations that are only partly aligned with the constructs or difficult to measure. The tasks also span multiple disciplines, limiting what the literature reveals about IS performance. Some findings suggest that DQ matters, but the literature mainly establishes that people respond to information about DQ rather than that DQ itself affects behavior. Addressing these gaps is necessary to establish DQ as a causal factor in IS research and inform practice. We now move to the second stage of the study and define a model based on the framework and findings above.

4.2. Methodological Development

To address limitations in previous attempts to study the causal impact of DQ, objective DQ and perceived quality must be independently and measurably manipulated. The following specification uses basic set notation to formalize these operational definitions. This notation explicitly defines the original dataset and the degree to which it is altered, enabling unambiguous DQ manipulation and facilitating comparison across

studies. The dimensional specifications build on established operationalizations where appropriate (e.g., [59,70]) while extending them through integration with the ontological model. This reduces ambiguity in the operationalization and manipulation of DQ in IS research.

Definitions: Let $I = 1, \ldots, n$ be the index set of the dataset. Let $\mathcal{E} = 1, \ldots, m$ be the set of entities contained within the dataset. Here, an entity denotes the underlying RW unit or event that a measurement corresponds to, whereas the dataset index is merely the position at which this measurement is stored. Let $\mathcal{V}$ denote the space of possible values, including the missing-value indicator N/A, while $\mathcal{F}$ represents possible syntactic representations. Let an entity-designation function be $g : I \to \mathcal{E}$, where g(i) identifies the entity e to which index i belongs e_i. Let $v : I \to \mathcal{V}$ be the value assigning function, where v(i) defines the value of the i-th entity. Then, let $f : \mathcal{V} \to \mathcal{F}$ be the syntax-assignment of a value, such as "five" or 5. A dataset X of size n is then defined as a function $X : I \to \mathcal{V} \times \mathcal{F}$, where $X(i) = x_i$ is the observed value of X at index i, given the value v_i and format f_i for the entity designated by g(i). For simplicity, it is said to be fully generated at the discrete time t. This will serve as the *correct dataset*, representing perfect DQ, such that each observed value corresponds to its true value, $I_X = I$, and $x_i \neq$ NA for all $i \in I$. A modified dataset X_e with an altered quality, generated from X, can be specified similarly: $X_e(i) : I_e \to \mathcal{V} \times \mathcal{F} : q_i, \quad I_e \subseteq I, \quad \mathcal{E}_e = g(I_e) \subseteq \mathcal{E}$. This generates at least two experimental conditions for assessing performance: participants making decisions with X and X_e, respectively. For simplicity, we exclude entity-class attributes and their indicators here, but the model can be extended to include them when their relationships are known. Such relationships can be defined in X and modified in X_e.

Accuracy: The specification must capture error magnitude while remaining conceptually separate from reliability (previous performance or the probability of being correct). We use functions for clarity and consistency when comparing operational definitions across studies. One way to induce errors ($X \to X_e$) is to consider the original distribution $\mathcal{L}(X)$ and assign new values to a percentage of X based on this distribution. These values represent incorrect but non-outlying instances at selected indices q_i. The error can be understood as mistakenly copying or entering a more common value instead of the correct one. To assess dataset-level accuracy without conflating it with completeness, accuracy should only consider elements that are present.

$$\text{Accuracy}(X_e) = 1 - \frac{|A|}{|I_e^+|}, \qquad A = \{i \in I_e^+ \mid q_i \neq x_i\}, \qquad I_e^+ = \{i \in I_e \mid q_i \neq \text{N/A}\}. \quad (1)$$

However, this conceptualization is not useful for tuple-level inaccuracy, as it would yield only 0 or 1. We must therefore also define error relative to the original value. Because this measure is highly domain-sensitive, percentage deviation from the true value provides one way to obtain comparable values across studies.

$$\text{Accuracy}(X_e) = 1 - \frac{\sum_{i \in I_e^+} |A_i|}{|I_e^+|}, \qquad A_i = \frac{x_i - q_i}{x_i}, \qquad I_e^+ = \{i \in I_e \mid q_i \neq \text{N/A}\}. \quad (2)$$

This formulation resolves prior specification issues [59]. It captures both over- and underestimation as errors while avoiding overlap with completeness. It is functionally equivalent to MAPE, a common forecasting loss function [76], and can generate fixed or varied error levels while remaining interpretable at both the tuple and dataset levels. Although MAPE fails when actual values are zero and is sensitive to extreme relative errors [77], control over X allows these issues to be managed experimentally.

DQI can distinguish knowledge from ignorance of DQ, enabling a 2×2 factorial design of objective and subjective DQ manipulations. Participants should be informed only about

the overall quality of the dataset, not individual items, specific errors, or error magnitude. Such detailed information could guide participants toward particular corrections rather than alter their perception of overall quality. Where possible, DQI should be conveyed naturally, for example, through a source indicating limited confidence in the dataset, evidence of quality controls, or other contextual cues signaling quality or its absence. This is preferable to explicitly presenting a quality metric, which may increase demand characteristics by signaling how participants are expected to interpret the data.

DQI is nevertheless imperfect for inducing perceived DQ because it informs rather than allows individuals to judge quality as they would in practice. The alternative, presenting information outside participants' expectations of reasonable values, places greater demands on their training and perceptual and cognitive processes, introducing multiple uncontrolled variables. These expectations are shaped by prior experience within the domain, expressable as $\mathrm{E}(X_e(t)) \sim \mathcal{L}(X_e(\tau < t))$, where τ denotes all previous experiences before testing. Because this process cannot be estimated precisely, and individual differences in expectations cannot be isolated, DQI provides a more controlled means of inducing a qualitative shift in perceived DQ. In some cases, natural experiments may be possible in which domain experts perform familiar tasks using datasets they normally use. A pilot study can then identify observations that fall within or outside participants' general expectations, allowing a qualitative shift in perception to be used to structure X and X_e.

Prior to the main experiment, pilot tests should assess whether the manipulation is reasonable. Following the task, an information-quality survey such as the AIMQ instrument [14] can assess the success and strength of the perceptual manipulation and support subsequent analyses.

Completeness: At the dataset level, completeness is the number of expected indices in the test set that have missing values. The modified dataset may also omit entire entities (missing rows/tuples), so entity completeness should likewise measure expected entities that are absent. Previous studies generally restricted missingness to selected attributes, which does not reflect how information typically goes missing [66–68]. Random missingness is more ecologically valid unless a specific attribute is theoretically relevant. Thus,

$$Completeness_v(X_e) = 1 - \frac{|C_v|}{|I|}, \quad C_v = \{i \in I \mid X_e(i) = \mathrm{N/A}\} \tag{3}$$

$$Completeness_e(X_e) = 1 - \frac{|C_e|}{|\mathcal{E}|}, \quad C_e = \{e \in \mathcal{E} \mid \nexists i \in I_e : g(i) = e\} \tag{4}$$

Strictly speaking, completeness may also refer to information sufficient for the task [13]. This raises difficult questions about what information is sufficient and what prior knowledge users possess, which are usually too complex to model. Researchers should instead compare performance using X and X_e in a domain where participants have no relevant prior knowledge, assuming that performance with the valid set X represents achievable performance. This is similar to previous work using benchmark performance as a reference [61,62].

Perceived completeness concerns whether participants recognize task-relevant information as present or absent. Unlike objective completeness, which can be manipulated by removing observations or values, perceived completeness depends on whether participants recognize this absence. Providing an unaware participant with a dataset with or without missing values will generally align perceived and objective completeness. However, perceived completeness can be manipulated in objectively incomplete datasets by retaining complete values for observed entities while omitting entities relevant to task performance. The dataset may therefore contain no visibly missing values while remaining incomplete at the entity level. Conversely, *perceived missingness* can be induced through natural cues

indicating absence, while the control condition provides no such cues. As with perceived accuracy, specific manipulations should be pilot-tested and assessed using a post-task information-quality survey [14].

Consistency: This dimension concerns whether multiple data points referring to the same entity agree. Although consistency is sometimes defined relative to external standards, an internal standard can be treated as a second data point referring to the same entity. To avoid overlap with accuracy, consistency should focus on agreement within the IS without reference to the true object. Thus, in experimental practice, consistency should concern data comparisons rather than correspondence with the represented object. This aligns with semiotic conceptualizations of DQ, which distinguish syntactic comparison from semantic representation [46]. In the ontological model, this corresponds to comparing IS states internally. Consistency therefore operates within the X_e space $\mathcal{V} \times \mathcal{F}$, such that tuples referring to the same entity agree or disagree regardless of whether they match the represented object. It can be divided into differences in data values (C_v) and differences in the format of the same entity value (C_f). Thus, data can be fully consistent without being accurate.

$$Consistency_v(X_e) = 1 - \frac{|C_v|}{|\mathcal{E}_e|}, \quad C_v = \{\, e \in \mathcal{E}_e \mid \exists i, j \in I_e : g(i) = g(j) \ \wedge \ v(i) \neq v(j) \,\} \tag{5}$$

$$Consistency_f(X_e) = 1 - \frac{|C_f|}{|\mathcal{E}_e|}, \quad C_f = \{\, e \in \mathcal{E}_e \mid \exists i, j \in I_e : g(i) = g(j) \ \wedge \ v(i) = v(j) \ \wedge \ f(i) \neq f(j) \,\} \tag{6}$$

Because of the above setup, value consistency (C_v) is less restrictive because it allows transformations that preserve the underlying value, whereas format consistency (C_f) is more restrictive and requires identical representations. If a total inconsistency metric is desired, the union $C_v \cup C_f$ captures all inconsistent entries arising from value or format differences. The two sets are disjoint because their value conditions are mutually exclusive: C_v requires different values, whereas C_f requires equal values.

The consistency setup requires assessing performance using either uncleaned or unmerged data within a single dataset, or multiple joined datasets containing disagreements. Format inconsistency can then be assessed using completion time as a separate dependent variable. *Perceived inconsistency* can be induced through value-consistent format differences, with the manipulation assessed using post-task surveys [14].

Timeliness: Task fit related to data age, and thus data currency, can be investigated by designing experimental tasks that receive frequent updates from a generative model and varying the dataset update frequency. This approach is similar to the limited previous work [70]. At time $t + 1$, the task is presented and dataset X is updated. A selected percentage of tuples (the timeliness condition) fails to update and retains its previous value from $X(t)$, generating X_e. Timeliness is therefore evaluated as

$$\text{Timeliness}(X_e) = 1 - \frac{|T|}{|I_e|}, \qquad T = \{i \in I_e \mid X_e(i, t+1) = X_e(i, t) \ \wedge \ X(i, t+1) \neq X(i, t)\}. \tag{7}$$

This provides a measure of dataset currency based on the proportion of values updated when required. As with completeness and some experimental accuracy studies, fitness-for-task can be benchmarked against performance using $X(t + 1)$ [61,62]. A simple experimental condition can use a generative model with a basic drift parameter for the entity and compare performance across different levels of timeliness. For *perceived timeliness*, participants can be presented with cues indicating temporal change and data currency, such as timestamps or update indicators, while delaying information to induce differences in perceived timeliness. Manipulation success can be assessed using a standard post-task information-quality survey [14].

4.3. Proof of Concept: Completeness Analysis

4.3.1. Setup

We now demonstrate the above framework using publicly available data. We use the well-known *Choices13k* dataset [78], which contains binary risky-choice lotteries with varied values and distributions for studying behavior under risk and ambiguity. Participants choose between two lotteries, with either complete or incomplete information about possible outcomes (Figure 2).

Relevance to this work: This dataset does not permit direct examination of accuracy, consistency, or timeliness as defined in our model, as all would receive perfect scores. The provided values and probabilities are treated as veridical representations of the intended experimental parameters and thus constitute the reference information set X defined by the experimenter rather than noisy data. However, the dataset's *ambiguity* parameter can be used to study information completeness. In particular, variation in outcome structure provides a basis for examining how incomplete representations affect decision behavior.

The present analysis does not re-evaluate the extensive literature on risky-choice prediction. Prior analyses of the *Choices13k* dataset have largely set aside ambiguity in favor of classical expected-utility or risk-based models, which assume known probabilities and stochastic outcomes [78,79]. Similarly, this study does not seek to compete with large-scale predictive models such as BEAST in predictive accuracy (e.g., CPC18 [80]). Instead, it focuses on interpretability and the construct-level effects of completeness to demonstrate the proposed methodological approach and its potential contribution to research on DQ and decision-making. Although the Ellsberg paradox (i.e., ambiguity aversion [80,81]) is relevant, the dataset does not include problems where lotteries A and B have identical structures but B has unknown probabilities. It therefore cannot test the paradox directly, although its effects may be inferred from the main effects of completeness.

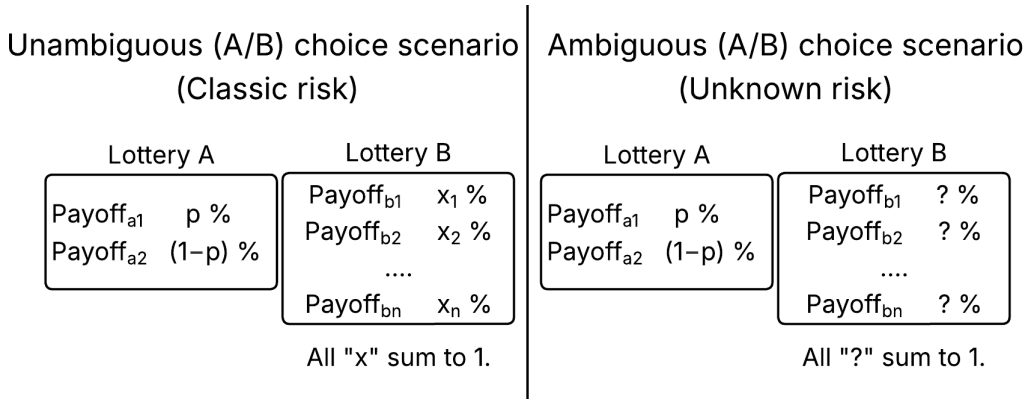


Figure 2. Risky choice setup for Choices13k dataset contrasting choice scenarios with lottery B having known payoff probabilities x_n with unknown payoff probabilities "?".

This proof of concept demonstrates how the proposed model can be implemented. As discussed, previous research provides an incomplete picture of completeness effects. Adequately powered studies have primarily examined qualitative differences, such as the presence or absence of arguments or details [66,68]. The main measurable studies have either been underpowered [59] or failed to reject the null for relevant outcomes such as preferred option, potentially because of weak manipulations [65,66,68]. Overall, prior research suggests that completeness may affect behavior and performance, motivating the following hypotheses.

H1. *In line with the Ellsberg paradox, more incomplete information increases uncertainty, inducing more ambiguity aversion and decreasing $P(B)$ at parity.*

H2. *Completeness moderates the relationship between value expectations and choice probability.*

Choices13k Experimental setup: The dataset examines risky choices between options (A and B) with different payoff structures [78]. It gained prominence for including 13,000 distinct problems, roughly 30 times more than previous datasets. A total of 14,711 participants were randomly selected to complete 20 problems, each repeated five times to estimate choice certainty. Participants were financially incentivized, with their eventual payout linked to their total gains across problems. They were also randomly assigned to receive feedback after each problem, with payoffs generated according to the problem structure. On average, 16.7 participants completed each problem, below the sample size considered desirable for stable problem-level estimates, and therefore introducing uncertainty. However, aggregating the five trials per participant provides more stable estimates of individual choices, while the large number of problems and participants provides sufficient information to estimate the aggregate choice function across levels of completeness. This does not eliminate uncertainty in individual-problem estimates but should be sufficient to estimate the shape of the choice function and provide broad insights.

The design can be described as a mixed-effects repeated-measures structure with one BS manipulation (feedback: yes/no) and multiple WS variations in value, structure, and completeness. In a conventional WS design, participants would complete identical or overlapping item sets, requiring careful counterbalancing to control for order and learning effects. In contrast, the dataset uses fully randomized item sets across participants, distributing potential order and learning effects across the large sample rather than systematically exposing participants to specific sets or orders. This design prioritizes population-level inference by treating observations as realizations of a stochastic decision process rather than repeated measurements of the same items within individuals. The focus therefore shifts from WS contrasts to estimating systematic effects of task features across a large and heterogeneous sample. The dataset is dominated by the completeness = 1 condition ($n = 11,759$), which was the original study's main focus. The incomplete conditions, which are of interest here, contain an average of 351 items each (minimum 67), providing sufficient observations for the present analysis.

Variable specification: Completeness is defined by the payoff values and probabilities presented to participants. $C_e = 0$ because all expected entities are present, while C_v depends on the amount of information hidden. Thus, complete conditions are identical to X and the incomplete conditions represents X_e as alterations to completeness. As discussed in Section 4.2, perceived and objective completeness coincide because only values, not entities, are omitted, and participants receive no information that challenges this perception.

By construction, lottery A contains four pieces of information from its two outcomes. Lottery B contains either $2n$ pieces of information when Ambiguity = 0 or n when all probabilities are missing (Ambiguity = 1). Completeness is therefore

$$Completeness_v(C13k) = 1 - \frac{|C_v|}{|I|} = 1 - \frac{Amb * n}{4 + 2n} \tag{8}$$

While completeness captures the number of options and their structure, participants may also respond to the values presented. Although not central to the analysis, we therefore specify four possible value-based strategies: (1) optimistic, (2) pessimistic, (3) equal-probability assumptions when B is unknown, and (4) expected value (EV). Table 5 summarizes these strategies, which are estimated in separate models to avoid multicollinearity. Here, v denotes possible payoffs. Strategy 3 draws on prior findings suggesting that individuals may assume equal probabilities under uncertainty or ignorance [55,57,82].

Table 5. Examined behavioral strategies no Choices13k data.

Strategy	Strategy Definition	Implementation
1. Optimistic	$\arg\max_{L\in\{A,B\}}(\max_{v\in L} v)$	$\max_{v\in A} v - \max_{v\in B} v$
2. Pessimistic	$\arg\max_{L\in\{A,B\}}(\min_{v\in L} v)$	$\min_{v\in A} v - \min_{v\in B} v$
3. Equal probability	$\arg\max_{L\in\{A,B\}}(\mathbb{E}[L]$ [1]$)$	$\mathbb{E}(A) - \sum(v(B))/n$
4. Expected value	$\arg\max_{L\in\{A,B\}}(\mathbb{E}[L])$	$\mathbb{E}(A) - \mathbb{E}(B)$

[1] Expected value under ambiguity is the assumption of equal weight.

We model the population-level probability of choosing lottery B as a function of value and structure, with a two-way interaction between value strategy and completeness. Feedback was included as a covariate to assess whether learning about choice outcomes affected overall choice.

We estimate a β-regression with a logit link, appropriate for proportional data and consistent with prospect-theoretic choice behavior [83]. For the 26 lotteries with B choice rates of exactly 0 or 1, we apply the Smithson–Verkuilen correction [84]. We compare model fit (AIC, BIC) and explanatory power (Ferrari & Cribari-Neta's pseudo-R^2 [85]), then assess completeness in the best-fitting model using odds ratios and interaction plots.

We also run separate analyses for complete and incomplete cases to account for potentially different value strategies when expected value cannot be evaluated. For comparability, the complete-information analysis uses the number of outcomes in B in place of completeness. Diagnostic tests, including Q-Q plots, posterior predictive checks, and homogeneity of variance tests, revealed no issues. The R (4.4.3; R Foundation for Statistical Computing, Vienna, Austria) code is provided in the Supplementary Material while the Choices13k data are publicly available.

4.3.2. Completeness Results

The best-performing strategy depended on completeness. Overall, expected value was the strongest predictor. When completeness = 1, comparing the expected value of A and B is the most rational choice strategy, which was predictive of choice behavior. In incomplete cases, however, behavior shifted toward the equal-probability assumption. Although the optimistic and pessimistic strategies were predictive, they were consistently outperformed, suggesting that they are not the primary drivers of choice. Choosing B also depended on completeness (main effect), and completeness moderated how utility was evaluated. Feedback had no significant effect.

Overall, the best-fitting models indicates substantial explanatory power: the full-sample expected value model ($R^2 = 0.37$), the incomplete-only equal-probability model ($R^2 = 0.30$), and the complete-only expected value model ($R^2 = 0.40$). Full model results are presented in Table 6.

Table 6. Full, incomplete-only, and complete-only sample regressions for P(B) over the value strategy, completeness, and feedback.

Model	Dataset	Wald OR [95% C.I]				Ferrari R^2	AIC	BIC
		Utility A-B [1]	Completeness [2]	Value | Completeness [2]	Feedback			
Optimistic	Full	0.99 *** [0.98, 0.99]	0.67 *** [0.58, 0.76]	1.01 *** [1.01, 1.01]	1.01 [0.97, 1.05]	0.04	−3922.04	−3876.52
	Incomplete	1.01 * [1.00, 1.02]	0.05 *** [0.03, 0.09]	0.98 *** [0.96, 0.99]	0.96 [0.88, 1.04]	0.12	−1053.14	−1017.50
	Complete	1.00 *** [1.00, 1.00]	1.03 *** [1.02, 1.04]	1.00 * [1.00, 1.00]	1.02 [0.97, 1.06]	0.02	−3007.09	−2962.86
Pessimistic	Full	0.98 *** [0.97, 0.98]	0.44 *** [0.39, 0.50]	1.01 *** [1.00, 1.01]	1.01 [0.98, 1.01]	0.10	−5035.03	−4989.51
	Incomplete	1.03 ** [1.01, 1.05]	0.06 *** [0.04, 0.11]	0.93 *** [0.91, 0.96]	0.97 [0.90, 1.06]	0.19	−1307.63	−1271.99
	Complete	0.98 *** [0.98, 0.98]	1.04 *** [1.03, 1.04]	1.00 *** [1.00, 1.00]	1.02 [0.98, 1.07]	0.09	−4049.99	−4005.75
Equal probability	Full	0.95 *** [0.95, 0.96]	0.82 ** [0.72, 0.94]	1.03 *** [1.03, 1.04]	1.02 [0.98, 1.06]	0.10	−5037.26	−4991.74
	Incomplete	1.11 *** [1.09, 1.13]	0.07 *** [0.04, 0.13]	0.82 *** [0.80, 0.85]	0.99 [0.92, 1.07]	0.30	−1786.86	−1751.22
	Complete	0.98 *** [0.97, 0.98]	1.04 *** [1.03, 1.04]	1.00 *** [1.00, 1.00]	1.02 [0.98, 1.07]	0.08	−3788.34	−3744.11
Expected value	Full	0.97 ** [0.96, 0.99]	0.48 *** [0.43, 0.53]	0.90 *** [0.89, 0.92]	1.02 [0.98, 1.05]	0.37	−10,916.33	−10,870.81
	Incomplete	0.99 [0.92, 1.06]	0.04 *** [0.02, 0.06]	0.89 * [0.81, 0.98]	0.97 [0.89, 1.05]	0.28	−1671.19	−1635.55
	Complete	0.89 *** [0.86, 0.92]	1.04 *** [1.03, 1.04]	1.00 *** [1.00, 1.01]	1.03 [0.99, 1.07]	0.40	−9665.30	−9621.06

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. [1] Variable depends on strategy by row as in Table 5. [2] For the complete-only dataset, the completeness variable is replaced by the number of options in lottery B.

For clarity of interpretation, which is particularly challenging with log-odds and interactions, we interpret the results using Figure 3. The graphs suggests that completeness significantly alters the decision process. As information becomes more incomplete, individuals become less certain that the dominant option is preferable, regardless of whether A or B dominates. The results also reveal distinct behavior under complete and incomplete information. The full-sample model does not capture this shift, as it is dominated by the complete cases, which constitute the majority of the sample. This is reflected in its improved explanatory power when incomplete cases are excluded. Instead, the incomplete cases better align with strategy (3). Thus, the results suggest a qualitative shift in strategy when information is both risky and incomplete, which the full-sample EV model cannot capture. These findings support H2. The superior performance of the equal-probability model in incomplete trials is also consistent with previous research on how people judge unknown probabilities [55,57,82].

In a supplementary analysis, we confirmed that the model predictions align with the Ellsberg paradox (H1). In the classical Ellsberg scenario (value and option difference = 0), the equal-probability model predicts $P(B) = 0.46$ under both feedback and no-feedback conditions, while the complete-information model predicts values closer to indifference ($P(B) = 0.49$ and 0.48, respectively). Because $P(B) < 0.5$ and the complete condition is closer to 0.5, these predictions are consistent with Ellsberg's ambiguity aversion, supporting H1 in this scenario.

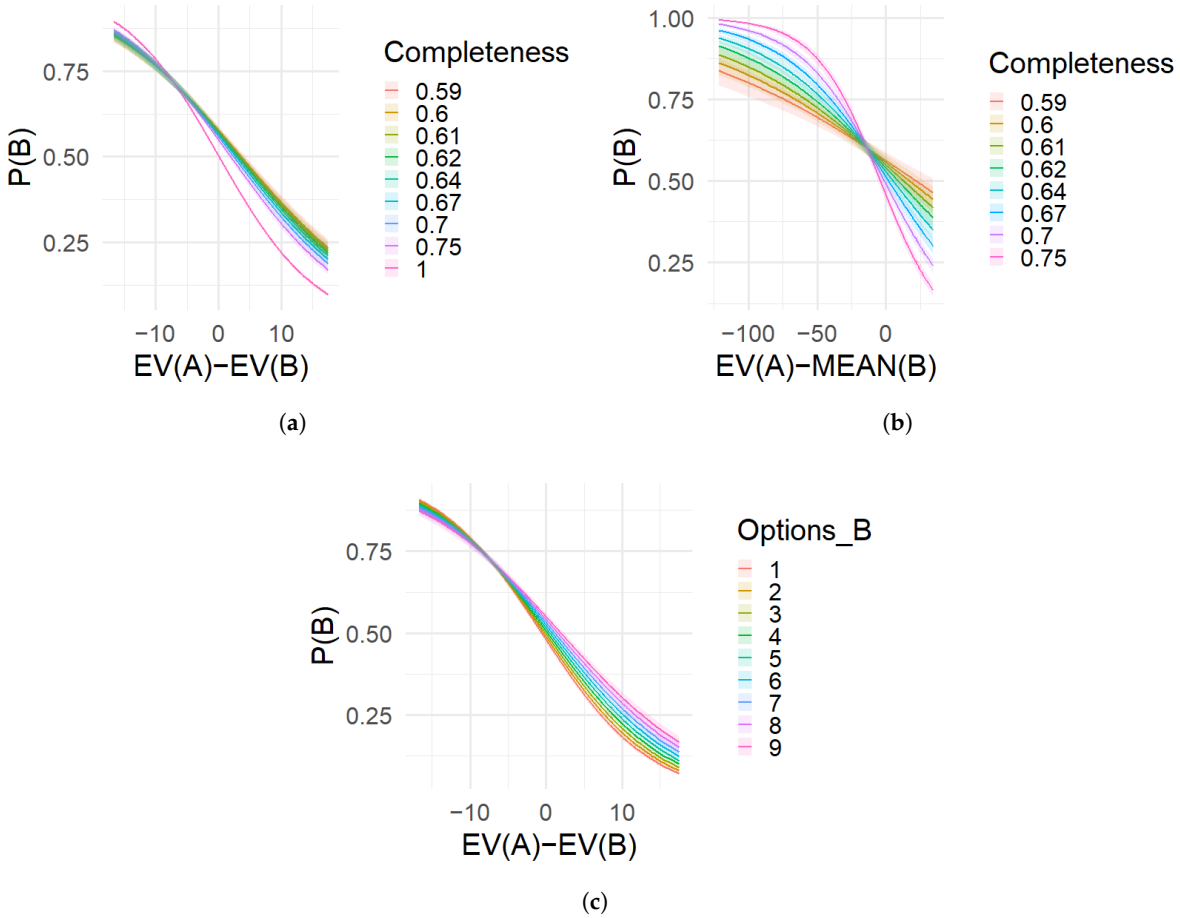


Figure 3. Beta-approximations of the choice function across the best-performing models in the respective datasets. (**a**) Expected value differences and completeness in the full sample. (**b**) Expected value difference to mean of B and completeness in the incomplete-only sample. (**c**) Expected value differences and B options in the complete-only sample.

Contrary to our expectation that greater incompleteness would increase aversion, this pattern reverses as completeness decreases, with preferences shifting toward B as the number of hidden $P(v_b)$ values increases. This is visible in Figure 3b, at $X = 0$. The same pattern appears in the raw data within a small range (± 5) around $EV(A) - Mean(B) = 0$. The standard Ellsberg setup (completeness = 0.75) points to a slight preference for A (mean = 0.48, median = 0.47), whereas cases with completeness < 0.75 indicate a preference for B (mean = 0.52, median = 0.54). More generally, the effect of completeness appears to increase with more extreme utility differences, consistent with the significant interaction between completeness and the relationship between $P(B)$ and value reported in Table 6.

Because similar effects were observed for the number of options and completeness in the complete-data sample, we conducted a post hoc test to assess the added value of completeness as a concept. We regressed completeness on the number of options and used the residuals as a predictor in a hierarchical model. Including residual completeness improved AIC by 35.1 (from -1757.5 to -1792.6), and a likelihood-ratio test confirmed a significant improvement in fit, $\chi^2(2) = 39.07$, $p < 0.001$. Thus, completeness explains additional variance in choice beyond the number of options in B, supporting its inclusion and indicating that preference for B does not simply increase with the number of outcomes it contains.

4.3.3. Discussion of Completeness Results

As discussed in the review, previous studies directly examining completeness have produced mixed results, ranging from non-significant effects to impacts on decision quality, although some were limited by small samples [59]. Related decision-making studies have examined varying levels of ambiguity in urn-choice tasks. One study found generally decreasing ambiguity aversion as ambiguity decreased in both adolescents and adults, consistent with our Ellsberg results [86]. Another found that greater ambiguity increased uncertainty, requiring larger utility differences to motivate the same choice, consistent with our finding that uncertainty has a stronger effect when utility differences are larger [87]. That study also identified an intra-individual amygdala response during ambiguous choices, consistent with the amygdala's role in emotional processing and the representation of positive and negative utility [87,88]. Finally, a study using partially hidden graphical distributions similarly found that ambiguity increased both its effect on choice and perceived risk, particularly with larger value differences [89].

Overall, our results are consistent with Ellsberg-paradox research (e.g., [90], replicating [81]) and studies manipulating missing information. However, the preference reversal as the number of outcomes under incompleteness increases appears to be novel, warranting further investigation.

4.4. General Discussion

The summative picture across these two phases is that experimental research on the causal impact of DQ remains underdeveloped. Existing studies often suffer from design limitations, including small samples and mismatched operationalizations, with mixed evidence of DQ effects. These limitations highlight the need for further research on DQ impacts, as well as potential trade-offs and interactions.

The proposed method can help resolve methodological ambiguities, facilitate comparison by mapping variables onto a common framework, and guide future applications. Its primary purpose is to clarify what is being studied when results are compared or synthesized, while explicitly distinguishing and manipulating objective and subjective DQ. Our proof of concept demonstrates the method's applicability to such research, but more direct and full application is warranted.

The present work highlights that the impact of DQ on decision-making remains poorly understood and requires greater attention. Consequently, the growing use of big data, open data, and public-sector data-driven initiatives proceeds without sufficient understanding of how specific DQ issues affect users and decision-makers [1–5]. Poor-quality data can also be difficult to identify before its consequences become apparent [17–19], potentially contributing to the persistent challenges faced by data-driven initiatives [91–94].

These findings suggest that ensuring intrinsic DQ before expanding data access may be preferable to exposing decision-makers to unknown risks. The completeness investigation identifies one such risk: reduced decisiveness and, in some scenarios, a preference for multiple unknowns over known risks.

The present work has several limitations. First, the review excludes non-causal empirical research, despite a substantial field of literature of well-designed, adequately powered studies that could provide additional insights [11,12]. Including this literature could broaden the evidence base, although it would also reintroduce the risk of causal misattribution discussed in the experimental theory section. Second, the review focuses on a single theoretical perspective [23] and excludes the common fitness-for-use view, which permits subjective appraisal of intrinsic dimensions. This was considered necessary because IS research contains numerous overlapping DQ concepts and definitions [44,45]. A measurable quality model therefore requires a clearly defined theoretical perspective; here, DQ is treated as objectively definable independent of context. Third, the review was conducted by a single author, so the analysis was not cross-checked by a second coder and may contain subjective judgments. Finally, the relatively small evidence base may reflect the field-specific terminology used in the review. A cross-disciplinary review incorporating alternative terms for the phenomena represented in the ontological model could reveal additional evidence. However, such terminology is difficult to identify systematically and may not map directly onto IS concepts. For example, ambiguity in risk analysis can encompass unknown distributions and unclear interpretation, overlapping with but not being ontologically identical to missing information and potentially reflecting different cognitive processes.

For the proof-of-concept analysis, stronger causal attribution could have been achieved using a BS design across completeness conditions. The average of 16.7 participants per problem also limits the precision of individual-problem estimates, although the large number of problems provides a reasonable approximation of the choice function. The Choices13k dataset also omits outcome probabilities rather than introducing random missingness, meaning that outcome values could also have been missing in a more natural design. Moreover, lottery A always contained two outcomes, which may have influenced results through its fixed structure and relative simplicity. For example, its expected value is easier to calculate than B's, potentially affecting choices through fatigue or information overload. Finally, experimental research base may have limited ecological validity. The dataset contains only a subset of management-adjacent tasks, while the proof of concept itself uses a highly abstract decision task.

Thus, further research should examine DQ in more ecologically valid settings while using designs that provide stronger causal evidence than existing approaches to operationalizing DQ. The methodological approach outlined here offers one pathway for doing so, while complementing rather than replacing correlational evidence. Further research should also address the identified limitations, test alternative operationalizations and contexts, and replicate and extend the findings of both the review and the proof-of-concept analysis.

5. Conclusions

DQ is essential for value generation, yet basic research on its mechanisms remains limited. This work moves beyond correlational evidence to examine the small body of causal research, revealing inconsistent and mismatched operationalizations, small sample sizes, weak controls and manipulations, and insufficient distinction between objective quality and subjective perception. Advancing the field requires robust methodological practices, including adequately powered experimental designs grounded in established DQ and methodological theory. We provide concrete design recommendations and demonstrate their applicability through a proof of concept. The empirical analysis identifies completeness as both a direct determinant of choice and a moderator of utility evaluation, with its negative effect on decisiveness becoming stronger under more extreme utility differences.

Future research should apply the model more fully and examine additional DQ and information-quality dimensions, such as accessibility, relevance, and reliability. As the present model considers only the micro-level case of an individual information user under controlled conditions, future work should examine DQ effects in organizational contexts shaped by data work, organizational cultures and structures, and institutional and group influences. Future research should also examine where causal and correlational evidence on DQ and decision-making converge or diverge. More ecologically valid experiments, cross-disciplinary evidence of adjacent constructs, and replication across contexts are needed to strengthen the evidence base. Addressing these directions can strengthen IS theory's predictive and explanatory power while increasing the practical value of DQ research for IS practice.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/informatics13080136/s1, File S1: Scoping Reviews (PRISMA-ScR) Checklist; File S2: R code.

Funding: This research received no external funding.

Data Availability Statement: The articles included in the review section of the study, and the discarded articles from the full reading phase of that review are openly available in Zenodo at https://doi.org/10.5281/zenodo.21909753. The data utilized in the study for the proof of concept are openly available in Github at https://github.com/jcpeterson/choices13k (accessed on 28 March 2026).

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BS	Between-subjects design
DM	Dataset manipulation
DQ	Data Quality
DQI	Data quality information
IS	Information systems
RW	Real-world
WS	Within-subjects design

References

1. Paliova, I. The Benefits of Digital Transformation for Governments. In *Proceedings of the Digital Transformation and Economic Recovery in the Post-Crisis Period: 10th Annual Conference, Sofia, Bulgaria, 18–20 November 2024*; Conference Proceedings; University of National and World Economy: Sofia, Bulgaria, 2024; pp. 156–165.
2. Park, S.; Gil-Garcia, J.R. Open Data Innovation: Visualizations and Process Redesign as a Way to Bridge the Transparency-Accountability Gap. *Gov. Inf. Q.* **2022**, *39*, 101456. [CrossRef]

3. Bertot, J.C.; Gorham, U.; Jaeger, P.T.; Sarin, L.C.; Choi, H. Big Data, Open Government and e-Government: Issues, Policies and Recommendations. *Inf. Polity* **2014**, *19*, 5–16. [CrossRef]
4. Irani, Z.; Abril, R.M.; Weerakkody, V.; Omar, A.; Sivarajah, U. The Impact of Legacy Systems on Digital Transformation in European Public Administration: Lesson Learned from a Multi Case Analysis. *Gov. Inf. Q.* **2023**, *40*, 101784. [CrossRef]
5. OECD. Digital Public Infrastructure: Government at a Glance 2025. 2025. Available online: https://www.oecd.org/en/publications/government-at-a-glance-2025_0efd0bcd-en/full-report/digital-public-infrastructure_1cee4220.html (accessed on 10 March 2026).
6. Yukhno, A. Digital Transformation: Exploring Big Data Governance in Public Administration. *Public Organ. Rev.* **2024**, *24*, 335–349. [CrossRef]
7. Pilshchikova, L.; Zuiderwijk, A.; Janssen, M. How Do Non-profit Open Data Intermediaries Enhance Open Data Usability? A Systematic Literature Review. In *Proceedings of the 18th International Symposium on Open Collaboration, Madrid Spain, 2022*; OpenSym '22; Association for Computing Machinery: New York, NY, USA, 2022; pp. 1–3. [CrossRef]
8. Romei, V.; Fleming, S. UK Inflation Overstated Due to Government Data Error, ONS Says. *Financial Times*, 5 June 2025.
9. Nikiforova, A. Definition and Evaluation of Data Quality: User-oriented Data Object-Driven Approach to Data Quality Assessment. *Balt. J. Mod. Comput.* **2020**, *8*, 391–432. [CrossRef]
10. Wang, J.; Liu, Y.; Li, P.; Lin, Z.; Sindakis, S.; Aggarwal, S. Overview of Data Quality: Examining the Dimensions, Antecedents, and Impacts of Data Quality. *J. Knowl. Econ.* **2023**, *15*, 1159–1178. [CrossRef] [PubMed]
11. Ghasemaghaei, M.; Ebrahimi, S.; Hassanein, K. Data Analytics Competency for Improving Firm Decision Making Performance. *J. Strateg. Inf. Syst.* **2018**, *27*, 101–113. [CrossRef]
12. Merriweather, J.; Lyytinen, K.; Aron, D.; Cauley, M.R. When Better Data Meets Better Design: How EHR Data Usability and System Usability Shape Physicians' Cognitive Load. *npj Digit. Med.* **2026**, *9*, 104. [CrossRef] [PubMed]
13. Wang, R.Y.; Strong, D.M. Beyond Accuracy: What Data Quality Means to Data Consumers. *J. Manag. Inf. Syst.* **1996**, *12*, 5–33. [CrossRef]
14. Lee, Y.W.; Strong, D.M.; Kahn, B.K.; Wang, R.Y. AIMQ: A Methodology for Information Quality Assessment. *Inf. Manag.* **2002**, *40*, 133–146. [CrossRef]
15. Raulin, M.L.; Graziano, A.M. Quasi-Experiments and Correlational Studies. In *Companion Encyclopedia of Psychology*; Routledge: Abington, UK, 1994.
16. Naser, M.Z. Causality and Causal Inference for Engineers: Beyond Correlation, Regression, Prediction and Artificial Intelligence. *WIREs Data Min. Knowl. Discov.* **2024**, *14*, e1533. [CrossRef]
17. Arazy, O.; Kopak, R.; Hadar, I. Heuristic Principles and Differential Judgments in the Quality Assessment of Information Quality. *J. Assoc. Inf. Syst.* **2017**, *18*, 403–432. [CrossRef]
18. Arazy, O.; Kopak, R. On the Measurability of Information Quality. *J. Am. Soc. Inf. Sci. Technol.* **2011**, *62*, 89–99. [CrossRef]
19. Klein, B.; Goodhue, D.L.; Davis, G.B. Can Humans Detect Errors in Data? Impact of Base Rates, Incentives, and Goals. *MIS Q.* **1997**, *21*, 169–194. [CrossRef]
20. Zhou, Z.; Fang, Y.; Straub, D. The Critical Need to Re-Balance Theory Building and Testing in the IS Field. *SSRN* 2023. [CrossRef]
21. Heeks, R.; Bailur, S. Analyzing E-Government Research: Perspectives, Philosophies, Theories, Methods, and Practice. *Gov. Inf. Q.* **2007**, *24*, 243–265. [CrossRef]
22. Aaltonen, A.; Alaimo, C.; Parmiggiani, E.; Stelmaszak, M.; Jarvenpaa, S.; Kallinikos, J.; Monteiro, E. What Is Missing from Research on Data in Information Systems? Insights from the Inaugural Workshop on Data Research. *Commun. Assoc. Inf. Syst.* **2023**, *53*, 475–490. [CrossRef]
23. Wand, Y.; Wang, R.Y. Anchoring Data Quality Dimensions in Ontological Foundations. *Commun. ACM* **1996**, *39*, 86–95. [CrossRef]
24. Declerck, J.; Kalra, D.; Vander Stichele, R.; Coorevits, P. Frameworks, Dimensions, Definitions of Aspects, and Assessment Methods for the Appraisal of Quality of Health Data for Secondary Use: Comprehensive Overview of Reviews. *JMIR Med. Inform.* **2024**, *12*, e51560. [CrossRef] [PubMed]
25. Hirschheim, R.; Klein, H. A Glorious and Not-So-Short History of the Information Systems Field. *J. Assoc. Inf. Syst.* **2012**, *13*, 188–235. [CrossRef]
26. Fink, L. Why and How Online Experiments Can Benefit Information Systems Research. *J. Assoc. Inf. Syst.* **2022**, *23*, 1333–1346. [CrossRef]
27. Douglas, H.; Magnus, P.D. State of the Field: Why Novel Prediction Matters. *Stud. Hist. Philos. Sci. Part A* **2013**, *44*, 580–589. [CrossRef]
28. Mortera, J.; Thompson, W.C. The Epistemic Value of Novel Predictive Success in Scientific and Criminal Investigations: A Bayesian Explanation. *Law Probab. Risk* **2025**, *24*, mgaf011. [CrossRef]
29. Nikiforova, A.; Bicevskis, J.; Bicevska, Z.; Oditis, I. Data Quality Model-Based Testing of Information Systems: The Use-Case of e-Scooters. In *Proceedings of the 2020 7th International Conference on Internet of Things: Systems, Management and Security, IOTSMS 2020, Paris, France, 14–16 December 2020*; IEEE: Piscataway, NJ, USA, 2020; pp. 1–8. [CrossRef]

30. Imai, K. *Quantitative Social Science: An Introduction*, 1st ed.; Princeton University Press: Princeton, NJ, USA, 2018.
31. Greenland, S. Randomization, Statistics, and Causal Inference. *Epidemiology* **1990**, *1*, 421–429. [CrossRef] [PubMed]
32. Greenwald, A.G. Within-Subjects Designs: To Use or Not to Use? *Psychol. Bull.* **1976**, *83*, 314–320. [CrossRef]
33. Cotton, J.W. *Analyzing Within-Subjects Experiments*; Lawrence Erlbaum Associates Publishers: Mahwah, NJ, USA, 1998; 352p.
34. Teo, T.S.H.; Wong, P.K. An Empirical Study of the Performance Impact of Computerization in the Retail Industry. *Omega* **1998**, *26*, 611–621. [CrossRef]
35. Austin, P.C.; Mamdani, M.M.; Van Walraven, C.; Tu, J.V. Quantifying the Impact of Survivor Treatment Bias in Observational Studies. *J. Eval. Clin. Pract.* **2006**, *12*, 601–612. [CrossRef] [PubMed]
36. Craighead, C.W.; Ketchen, D.J., Jr.; Dunn, K.S.; Hult, G.T.M. Addressing Common Method Variance: Guidelines for Survey Research on Information Technology, Operations, and Supply Chain Management. *IEEE Trans. Eng. Manag.* **2011**, *58*, 578–588. [CrossRef]
37. Islam, M.R. Sample Size and Its Role in Central Limit Theorem (CLT). *Comput. Appl. Math. J.* **2018**, *4*, 1–7.
38. Kwak, S.G.; Kim, J.H. Central Limit Theorem: The Cornerstone of Modern Statistics. *Korean J. Anesthesiol.* **2017**, *70*, 144–156. [CrossRef] [PubMed]
39. Aziz, S.; Nayem, H.M.; Kibria, B.G. Sample Size Requirements for the Central Limit Theorem for Skewed Distributions: A Simulation Study. *J. Stat. Model. Anal. (JOSMA)* **2025**, *7*. [CrossRef]
40. Aaltonen, A.; Stelmaszak, M. The Performative Production of Trace Data in Knowledge Work. *Inf. Syst. Res.* **2024**, *35*, 1448–1462. [CrossRef]
41. Robinson, K.; Cummings, E.; Barraclough, S.; iedema, R. Creating Order from Disorder in Health Data and Health Information Systems. A Problematic of Invisibility of the 'Truth Makers'. *Scand. J. Inf. Syst.* **2025**, *37*, 12. [CrossRef]
42. Aaltonen, A.; Alaimo, C.; Kallinikos, J. The Making of Data Commodities: Data Analytics as an Embedded Process. *J. Manag. Inf. Syst.* **2021**, *38*, 401–429. [CrossRef]
43. Mikalsen, M.; Monteiro, E. Acting with Inherently Uncertain Data: Practices of Data-Centric Knowing. *J. Assoc. Inf. Syst.* **2021**, *22*, 1715–1735. [CrossRef]
44. Madnick, S.E.; Wang, R.Y.; Lee, Y.W.; Zhu, H. Overview and Framework for Data and Information Quality Research. *J. Data Inf. Qual.* **2009**, *1*, 2:1–2:22. [CrossRef]
45. Sahi, L.; Laborde, R.; Kandi, M.A.; Sibilla, M.; Macilotti, G.; Abdelmalek, B.; Ferreira, A. Towards Reliable Collaborative Data Processing Ecosystems: Survey on Data Quality Criteria. In *Proceedings of the 2023 IEEE 22nd International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), Exeter, UK, 2023*; IEEE: Piscataway, NJ, USA, 2023; pp. 2456–2464. [CrossRef]
46. Price, R.; Shanks, G. A Semiotic Information Quality Framework: Development and Comparative Analysis. In *Enacting Research Methods in Information Systems: Volume 3*; Willcocks, L.P., Sauer, C., Lacity, M.C., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 219–250. [CrossRef]
47. Zhang, L.; Jeong, D.; Lee, S. Data Quality Management in the Internet of Things. *Sensors* **2021**, *21*, 5834. [CrossRef] [PubMed]
48. Alipour, J.; Ahmadi, M. Dimensions and Assessment Methods of Data Quality in Health Information Systems. *Acta Medica Mediterr.* **2017**, *33*, 313–320. [CrossRef]
49. Gusenbauer, M.; Haddaway, N.R. Which Academic Search Systems Are Suitable for Systematic Reviews or Meta-Analyses? Evaluating Retrieval Qualities of Google Scholar, PubMed, and 26 Other Resources. *Res. Synth. Methods* **2020**, *11*, 181–217. [CrossRef] [PubMed]
50. Kitchenham, B.; Charters, S. *Guidelines for Performing Systematic Literature Reviews in Software Engineering*; Technical Report 2.3; Keele University: Keele, UK, 2007.
51. Tricco, A.C.; Lillie, E.; Zarin, W.; O'Brien, K.K.; Colquhoun, H.; Levac, D.; Moher, D.; Peters, M.D.; Horsley, T.; Weeks, L.; et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann. Intern. Med.* **2018**, *169*, 467–473. [CrossRef] [PubMed]
52. Kraus, S.; Breier, M.; Lim, W.M.; Dabić, M.; Kumar, S.; Kanbach, D.; Mukherjee, D.; Corvello, V.; Piñeiro-Chousa, J.; Liguori, E.; et al. Literature Reviews as Independent Studies: Guidelines for Academic Practice. *Rev. Manag. Sci.* **2022**, *16*, 2577–2595. [CrossRef]
53. Webster, J.; Watson, R.T. Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Q.* **2002**, *26*, xiii–xxiii. [CrossRef]
54. Reader, K.; Kleiner, B. Personal Computing in the 1990s. *J. Int. Inf. Manag.* **1992**, *1*, 8. [CrossRef]
55. Kelly, M.O.; Budescu, D.V.; Dhami, M.; Mandel, D.R. The Effect of Source Reliability and Information Credibility on Judgments of Information Quality in Intelligence Analysis. *Judgm. Decis. Mak.* **2025**, *20*, e36. [CrossRef]
56. Budescu, D.V.; Rantilla, A.K.; Yu, H.T.; Karelitz, T.M. The Effects of Asymmetry among Advisors on the Aggregation of Their Opinions. *Organ. Behav. Hum. Decis. Process.* **2003**, *90*, 178–194. [CrossRef]

57. Mandel, D.R.; Irwin, D.; Dhami, M.K.; Budescu, D.V. Meta-Informational Cue Inconsistency and Judgment of Information Accuracy: Spotlight on Intelligence Analysis. *J. Behav. Decis. Mak.* **2023**, *36*, 14. [CrossRef]
58. Fisher, C.W.; Chengalur-Smith, I.; Ballou, D.P. The Impact of Experience and Time on the Use of Data Quality Information in Decision Making. *Inf. Syst. Res.* **2003**, *14*, 170–188 [CrossRef]
59. Ge, M.; Helfert, M. Impact of Information Quality on Supply Chain Decisions. *J. Comput. Inf. Syst.* **2013**, *53*, 59–67. [CrossRef]
60. Shikano, S.; Stoffel, M.F.; Tepe, M. Information Accuracy in Legislative Oversight: Theoretical Implications and Experimental Evidence. *Ration. Soc.* **2017**, *29*, 226–254. [CrossRef]
61. Shankaranarayanan, G.; Zhu, B. Data Quality Metadata and Decision Making. In *Proceedings of the Annual Hawaii International Conference on System Sciences*; IEEE Computer Society; IEEE: Piscataway, NJ, USA, 2012; pp. 1434–1443. [CrossRef]
62. Moges, H.T.; Vlasselaer, V.; Lemahieu, W.; Baesens, B. Determining the Use of Data Quality Metadata (DQM) for Decision Making Purposes and Its Impact on Decision Outcomes—An Exploratory Study. *Decis. Support Syst.* **2016**, *83*, 32–46. [CrossRef]
63. Chengalur-Smith, I.; Ballou, D.P.; Pazer, H.L. The Impact of Data Quality Information on Decision Making: An Exploratory Analysis. *IEEE Trans. Knowl. Data Eng.* **1999**, *11*, 853–864. [CrossRef]
64. Price, R.J.; Shanks, G. The Impact of Data Quality Tags on Decision-Making Outcomes and Process. *J. Assoc. Inf. Syst.* **2011**, *12*, 323–346. [CrossRef]
65. Sah, S.; Feiler, D. Conflict of Interest Disclosure with High-Quality Advice: The Disclosure Penalty and the Altruistic Signal. *Psychol. Public Policy Law* **2020**, *26*, 88–104. [CrossRef]
66. Kadous, K.; Mercer, M.; Zhou, Y. Why Do Investors Rely on Low-Quality Investment Advice? Experimental Evidence from Social Media Platforms. *Behav. Res. Account.* **2025**, *37*, 97–115. [CrossRef]
67. Ahituv, N.; Igbaria, M.; Sella, A.V. The Effects of Time Pressure and Completeness of Information on Decision Making. *J. Manag. Inf. Syst.* **1998**, *15*, 153–172. [CrossRef]
68. Nurse, J.R.; Agrafiotis, I.; Creese, S.; Goldsmith, M.; Lamberts, K. Building Confidence in Information-Trustworthiness Metrics for Decision Support. In *Proceedings of the 2013 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (Trustcom 2013)*; IEEE: Piscataway, NJ, USA, 2013; pp. 535–543. [CrossRef]
69. Martire, K.A.; Edmond, G.; Navarro, D. Exploring Juror Evaluations of Expert Opinions Using the Expert Persuasion Expectancy Framework. *Leg. Criminol. Psychol.* **2020**, *25*, 90–110. [CrossRef]
70. Blake, R.; Mangiameli, P. The Effects and Interactions of Data Quality and Problem Complexity on Classification. *J. Data Inf. Qual.* **2011**, *2*, 1–28. [CrossRef]
71. Timmerman, Y.; Nasfi, R.; De Tré, G.; Pattyn, F.; Bronselaer, A. Cost-Based Analysis of the Impact of Data Completeness and Representational Consistency. *Decis. Support Syst.* **2023**, *175*, 114044. [CrossRef]
72. Jung, W.; Ryan, T.; Olfman, L.; Park, Y.T. An Experimental Study of the Effects of Representational Data Quality on Decision Performance. In *Proceedings of the AMCIS 2005 Proceedings, Omaha, 2005; Vol. 4, Association for Information Systems—11th Americas Conference on Information Systems, AMCIS 2005: A Conference on a Human Scale*; Association for Information Systems (AIS): Atlanta, Georgia, USA, 2005; pp. 2044–2052.
73. Letzring, T.D.; Wells, S.M.; Funder, D.C. Information Quantity and Quality Affect the Realistic Accuracy of Personality Judgment. *J. Personal. Soc. Psychol.* **2006**, *91*, 111–123. [CrossRef] [PubMed]
74. Cochran, R.N.; vanDellen, M.R.; Haas, B.W. Getting to Know You: Associations between Judge and Target Personality with Personality Judgment Accuracy during a Dyadic Social Interaction Task. *Personal. Individ. Differ.* **2019**, *142*, 139–144. [CrossRef]
75. Liao, H.L.; Huang, Z.Y.; Liu, S.H. The Effects of Negative Online Reviews on Consumer Perception, Attitude and Purchase Intention: Experimental Investigation of the Amount, Quality, and Presentation Order of eWOM. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* **2021**, *20*, 51. [CrossRef]
76. Jadon, A.; Patil, A.; Jadon, S. A Comprehensive Survey of Regression-Based Loss Functions for Time Series Forecasting. In *Proceedings of the Data Management, Analytics and Innovation*; Sharma, N., Goje, A.C., Chakrabarti, A., Bruckstein, A.M., Eds.; Springer: Singapore, 2024; pp. 117–147. [CrossRef]
77. Swanson, D.A.; Tayman, J.; Bryan, T.M. MAPE-R: A Rescaled Measure of Accuracy for Cross-Sectional Subnational Population Forecasts. *J. Popul. Res.* **2011**, *28*, 225–243. [CrossRef]
78. Peterson, J.C.; Bourgin, D.D.; Agrawal, M.; Reichman, D.; Griffiths, T.L. Using Large-Scale Experiments and Machine Learning to Discover Theories of Human Decision-Making. *Science* **2021**, *372*, 1209–1214. [CrossRef] [PubMed]
79. Mullainathan, S.; Rambachan, A. *From Predictive Algorithms to Automatic Generation of Anomalies*; National Bureau of Economic Research: Cambridge, MA, USA, 2024. [CrossRef]
80. Erev, I.; Ert, E.; Plonsky, O.; Cohen, D.; Cohen, O. From Anomalies to Forecasts: Toward a Descriptive Model of Decisions under Risk, under Ambiguity, and from Experience. *Psychol. Rev.* **2017**, *124*, 369–409. [CrossRef] [PubMed]
81. Ellsberg, D. Risk, Ambiguity, and the Savage Axioms. *Q. J. Econ.* **1961**, *75*, 643–669. [CrossRef]
82. Camerer, C.; Weber, M. Recent Developments in Modeling Preferences: Uncertainty and Ambiguity. *J. Risk Uncertain.* **1992**, *5*, 325–370. [CrossRef]

83. Kahneman, D.; Tversky, A. Prospect Theory: An Analysis of Decision under Risk. *Econometrica* **1979**, *47*, 263–291. [CrossRef]
84. Smithson, M.; Verkuilen, J. A Better Lemon Squeezer? Maximum-likelihood Regression with Beta-Distributed Dependent Variables. *Psychol. Methods* **2006**, *11*, 54–71. [CrossRef] [PubMed]
85. Ferrari, S.; Cribari-Neto, F. Beta Regression for Modelling Rates and Proportions. *J. Appl. Stat.* **2004**, *31*, 799–815. [CrossRef]
86. Osmont, A.; Cassotti, M. Development of Ambiguity Aversion from Early Adolescence to Adulthood: New Insights from the Ellsberg Paradox. *Int. J. Behav. Dev.* **2023**, *47*, 47–58. [CrossRef]
87. Levy, I.; Snell, J.; Nelson, A.J.; Rustichini, A.; Glimcher, P.W. Neural Representation of Subjective Value Under Risk and Ambiguity. *J. Neurophysiol.* **2010**, *103*, 1036–1047. [CrossRef] [PubMed]
88. Phelps, E.A. Emotion and Cognition: Insights from Studies of the Human Amygdala. *Annu. Rev. Psychol.* **2006**, *57*, 27–53. [CrossRef] [PubMed]
89. Blankenstein, N.E.; van Duijvenvoorde, A.C.K. Neural Tracking of Subjective Value under Riskand Ambiguity in Adolescence. *Cogn. Affect. Behav. Neurosci.* **2019**, *19*, 1364–1378. [CrossRef] [PubMed]
90. Fairley, K.; Sanfey, A.G. The Role of Demographics on Adolescents' Preferences for Risk, Ambiguity, and Prudence. *J. Econ. Behav. Organ.* **2020**, *179*, 784–796. [CrossRef]
91. Dada, D. The Failure of E-Government in Developing Countries: A Literature Review. *Electron. J. Inf. Syst. Dev. Ctries.* **2006**, *26*, 1–10. [CrossRef]
92. Choi, T.; Chandler, S.M. Knowledge Vacuum: An Organizational Learning Dynamic of How e-Government Innovations Fail. *Gov. Inf. Q.* **2020**, *37*, 101416. [CrossRef]
93. Budzier, A.; Flyvbjerg, B. Overspend? Late? Failure? What the Data Say About IT Project Risk in the Public Sector. *arXiv* **2013**. arXiv:1304.4525v3.
94. van Ooijen, C.; Ubaldi, B.; Welby, B. A Data-driven Public Sector: Enabling the Strategic Use of Data for Productive, Inclusive and Trustworthy Governance. In *OECD Working Papers on Public Governance*; OECD Publishing: Paris, France, 2019. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.