behavioral sciences

MDPI

Article

Managerial Judgment Under Ambiguity: Subjective Value, Confidence Calibration, and a Behaviorally Calibrated Fuzzy Model

Filiz Mizrak [1,*], Hatice Gokce Demirel [2] and Turhan Karakaya [3]

[1] Management Information Systems, Atlas University, Istanbul 34408, Turkey
[2] Business Administration, Fırat University, Elazig 23119, Turkey; demirel@firat.edu.tr
[3] Industrial Engineering, Doğuş University, Istanbul 34680, Turkey
* Correspondence: filiz.mizrak@atlas.edu.tr

Abstract

This study examines how ambiguity shapes managerial subjective value and decision commitment and evaluates a Behaviorally Calibrated Fuzzy Subjective-Value (BCFSV) representation of ambiguity responses. In a repeated-measures 2 × 2 experiment crossing risk versus ambiguity with gain versus loss framing, 525 working adults completed 12 managerial decision scenarios, yielding 6300 trial-level observations. An independent pilot ($N = 42$; 504 trials) was used to calibrate the fuzzy hesitation/ambiguity-response term (π) against normalized perceived ambiguity using ambiguity exposure, ambiguity width, intolerance of uncertainty, and their interaction. In the main sample, ambiguity increased perceived ambiguity and reduced the pre-choice subjective value of the uncertain option; the statistical indirect effect through perceived ambiguity was −0.405 (95% bootstrap CI [−0.439, −0.373]). Intolerance of uncertainty strengthened the ambiguity–perceived ambiguity relationship, and framing moderated the ambiguity–valuation relationship. The highest pre-choice subjective value across the two options predicted decision commitment and carried a significant indirect association between perceived ambiguity and commitment (−1.584, 95% bootstrap CI [−1.784, −1.391]). Decision confidence also predicted commitment, with a stronger confidence–commitment association among participants showing poorer Raw Brier calibration ($b = 0.874$, 95% CI [0.630, 1.119]). The pilot-calibrated π mapping generalized to the main sample ($r = 0.620$), but adding the BCFSV transformation to a same-information benchmark produced no detectable incremental predictive gain for choice or commitment. The findings indicate that ambiguity appraisal, individual differences, and framing shape valuation, while the confidence–commitment association was stronger among participants with poorer confidence calibration. BCFSV thus offers an interpretable, empirically calibrated representation of uncertainty rather than a predictively superior model.

Keywords: managerial judgment; ambiguity; subjective value; confidence calibration; decision commitment; fuzzy hesitation

Check for updates

Academic Editors: Petko Kusev and Mimi Fakhri

Received: 5 September 2026
Revised: 20 October 2026
Accepted: 24 October 2026
Published: 28 October 2026

Copyright: © 2026 by the authors. 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.

1. Introduction

Managers rarely make important decisions with complete information. Choices about market entry, technology investment, suppliers, innovation, capacity, and organizational change often have to be made while future outcomes remain uncertain. Importantly, not all uncertainty is the same. Under risk, the possible outcomes are uncertain but their

probabilities are known with reasonable precision. Under ambiguity, the probabilities themselves may be incomplete, disputed, or expressed as ranges. A manager may therefore react differently to a known 30% probability of failure than to an estimate suggesting that the probability could lie anywhere between 15% and 45%. Research consistently distinguishes these two situations and shows that responses to ambiguity depend on its degree and structure rather than simply on whether ambiguity is present (Bühren et al., 2023; Klingebiel & Zhu, 2023; J. Zhang, 2022; Aydogan et al., 2026).

One reason ambiguity matters is that managers do not evaluate decisions only through objective probabilities and expected payoffs. They also form a subjective value—a personal assessment of how attractive an option appears in a particular context. This evaluation depends on the information available, the way the decision is framed, and characteristics of the decision maker (Zheng et al., 2026). When probabilities are imprecise, expected-value reasoning also becomes less straightforward (Jiroušek & Kratochvíl, 2020). Consequently, two managers may receive the same information yet value the same opportunity differently. One explanation is perceived ambiguity: the extent to which information feels unclear, difficult to evaluate, or unreliable. Individual differences also matter. Managers who are less comfortable with uncertainty may experience the same ambiguous information as more problematic, consistent with research on intolerance of uncertainty, indecisiveness, and information-seeking behavior (Appel et al., 2024; Bartoszek et al., 2022; Cobos et al., 2024). Although ambiguity has received increasing attention, its translation from objective uncertainty into perceived ambiguity, subjective value, and eventual commitment remains less fully developed in managerial decision research.

A related issue is decision confidence. Managers often have to act before uncertainty disappears, making confidence an important part of commitment and action. Yet feeling confident is not the same as being correct. Confidence can guide effort, information use, and willingness to act, but it is also an imperfect signal of judgment quality (Aguilar-Lleyda & de Gardelle, 2021; Balsdon & Philiastides, 2024; Fleming, 2024; Boundy-Singer et al., 2023). Research on managerial overconfidence likewise shows that excessive certainty can shape strategic risk taking, investment, resilience, and performance (Brunzel, 2021; Burkhard et al., 2023; Kunz & Sonnenholzner, 2023). Recent studies further suggest that confidence may increase during the decision process even when decision quality does not improve (Boyle et al., 2025; Boyle & Nye, 2026). This makes confidence calibration especially important: a highly confident manager may be accurate, inaccurate, or simply more certain than the available evidence justifies. Distinguishing confidence from calibration therefore offers a clearer way to understand why some managers commit strongly to decisions even when their confidence is poorly supported by performance.

A further challenge is methodological. Fuzzy approaches are useful for representing incomplete and imprecise information, and they have increasingly been combined with prospect-based and behavioral decision models (Tian et al., 2022; Pan et al., 2024; Wu et al., 2024). Recent studies have also introduced more interpretable and personalized fuzzy structures (García-Zamora et al., 2024, 2026; Dutta et al., 2026). However, personalization is often based on predefined linguistic mappings or judgments elicited directly from decision makers or experts. A remaining opportunity is to connect fuzzy uncertainty more closely to observed behavioral appraisal. In particular, a fuzzy hesitation/ambiguity-response term can be empirically calibrated from pre-choice ambiguity judgments and allowed to vary with objective ambiguity, ambiguity width, and stable individual differences such as intolerance of uncertainty. Confidence calibration can then be examined separately as a metacognitive moderator of post-choice commitment rather than being folded into the primary hesitation equation.

Against this background, the present study examines how objective ambiguity is related to perceived ambiguity, how that perception is associated with subjective value and decision commitment, and how intolerance of uncertainty, gain/loss framing, confidence, and confidence calibration condition these relationships. Using a repeated-measures managerial decision experiment, the study develops a Behaviorally Calibrated Fuzzy Subjective-Value (BCFSV) model in which the hesitation/ambiguity-response term is calibrated in an independent pilot from normalized perceived ambiguity as a function of ambiguity exposure, ambiguity width, intolerance of uncertainty, and their interaction. Confidence calibration is analyzed separately as a moderator of the confidence–commitment relationship. The study contributes in three ways. First, it evaluates a behavioral appraisal pathway linking ambiguity with managerial valuation and commitment. Second, it distinguishes expressed confidence from confidence calibration when examining commitment. Third, it links independently observed ambiguity appraisals to fuzzy uncertainty representation and evaluates whether the BCFSV transformation adds predictive information beyond models containing the same underlying inputs. In doing so, the study treats uncertainty not simply as a feature of the decision itself, but as an interaction between the information environment and the manager interpreting it.

2. Literature Review and Hypothesis Development

2.1. Managerial Judgment Under Risk and Ambiguity

Managerial decisions are rarely made with complete information. Choices concerning market entry, investment, innovation, strategic partnerships, technology adoption, and project continuation frequently require managers to commit resources before important outcomes are known. Yet not all forms of uncertainty are equivalent. A distinction that is particularly important for managerial judgment is the difference between risk and ambiguity. Under risk, possible outcomes remain uncertain but their probabilities are known or can be estimated with reasonable precision. Under ambiguity, the probability structure itself is incomplete, imprecise, or contested. This distinction matters because decision makers do not respond to unknown probabilities in the same way that they respond to known risks (Bühren et al., 2023; Klingebiel & Zhu, 2023; J. Zhang, 2022).

Recent work has moved away from treating ambiguity as a simple binary condition. Klingebiel and Zhu (2023), for example, show that reactions depend on the degree of ambiguity, while Aydogan et al. (2026) demonstrate that ambiguity aversion is more heterogeneous than a general preference for known probabilities would suggest. Experimental evidence similarly indicates that uncertainty does not automatically translate into greater risk aversion (Leder et al., 2024). People may respond differently depending on how uncertainty is structured, what information is available, and how much control they believe they have over the decision. Randomization advice, information structures, and alternative axiomatizations of ambiguity further illustrate that behavioral responses cannot be reduced to a single stable ambiguity-aversion parameter (Chen, 2026; Kuzmics et al., 2024).

This distinction is especially consequential in management. An investment with a known 30% probability of failure presents a different psychological problem from an investment for which estimates of failure range between 10% and 50%. In the second situation, the manager is not merely evaluating an unfavorable probability; the manager must also decide how much confidence to place in the probability information itself. Studies of professional judgment show that ambiguity can affect choices even when decision makers operate in consequential and expertise-intensive environments (Almazrouei et al., 2025). Thus, managerial ambiguity should not be treated simply as "more risk."

Subjective-value approaches provide a useful way to explain why this distinction changes choice. Subjective value refers to the value that a decision maker psychologically

assigns to an option rather than the option's objective monetary or statistical value. Recent synthesis suggests that subjective value under uncertainty is constructed through the interaction of outcome characteristics, probability information, affect, context, and individual differences (Zheng et al., 2026). Earlier formal work likewise shows that expected value becomes more difficult to define when beliefs themselves are ambiguous rather than probabilistically precise (Jiroušek & Kratochvíl, 2020).

For managerial decisions, this implies that an objectively attractive opportunity may lose subjective value simply because the probability information supporting it is unclear. Importantly, this does not mean that ambiguity always causes avoidance. Instead, ambiguity changes the psychological representation of the alternative, making the eventual valuation more dependent on the decision maker and the context. Nevertheless, when expected outcomes are held comparable, ambiguous probability information should generally make valuation more cautious because the decision maker must price both outcome uncertainty and uncertainty concerning the probability estimate itself.

H1. *Managerial alternatives presented under ambiguity will receive lower subjective-value evaluations than equivalent alternatives presented under measurable risk.*

Objective ambiguity, however, is unlikely to influence subjective value entirely through direct calculation. Decision makers first interpret whether the information feels vague, unreliable, incomplete, or difficult to evaluate. This perceived ambiguity provides the psychological bridge between the informational environment and subjective valuation. The recent literature on subjective value emphasizes precisely this constructive nature of valuation, while studies of context demonstrate that value representations shift according to the way options and information are presented (Soukupova et al., 2026; Zheng et al., 2026). Accordingly, two formally equivalent ambiguous situations may produce different valuations if they are not experienced as equally ambiguous.

H2. *Objective ambiguity will show a significant statistical indirect effect on subjective value through perceived ambiguity: ambiguity will increase perceived ambiguity, which will in turn be associated with lower subjective value.*

2.2. Intolerance of Uncertainty as an Individual-Difference Boundary Condition

The effect of ambiguity should not be assumed to be uniform across managers. Some people can continue evaluating alternatives despite incomplete information, whereas others experience uncertainty itself as difficult or aversive. Intolerance of uncertainty captures this individual tendency to react negatively when important outcomes, probabilities, or consequences cannot be known with confidence.

Recent evidence connects intolerance of uncertainty with indecisiveness, information seeking, safety behavior, and other attempts to reduce uncertainty (Appel et al., 2024; Bartoszek et al., 2022). Zhao et al. (2022), for example, show that intolerance of uncertainty can influence behavior through perceived risk, suggesting that individual differences operate partly by altering how uncertain situations are interpreted rather than simply changing preferences after risk has already been assessed. Related experimental work also connects intolerance of uncertainty to exploration and exploitation decisions, although the effects are not uniformly strong across computational parameters (Krypotos et al., 2022).

This qualification is important. Intolerance of uncertainty should not be treated as a universal predictor of every risky choice. Cobos et al. (2024) failed to find a strong relationship between intolerance of uncertainty and decisions involving delayed or probabilistic rewards, illustrating that reactions to uncertainty depend on the nature of the decision problem. That finding actually strengthens the rationale for distinguishing ambiguity

from ordinary probability-based risk. An individual may tolerate a clearly specified 60% probability while reacting much more strongly when the same probability is described as uncertain or disputed.

Loss-related ambiguity may further intensify these responses. Hu et al. (2023) show that ambiguity and loss processing involve distinguishable behavioral and neural mechanisms, reinforcing the idea that the subjective experience of ambiguity cannot be inferred from objective risk alone. In managerial settings, intolerance of uncertainty should therefore influence the translation of objective ambiguity into perceived ambiguity. A manager with low intolerance of uncertainty may regard a wide probability interval as manageable strategic uncertainty, whereas a manager with high intolerance of uncertainty may interpret the same information as a serious obstacle to commitment.

H3. *Intolerance of uncertainty will moderate the relationship between objective ambiguity and perceived ambiguity, such that the positive effect of objective ambiguity on perceived ambiguity will be stronger at higher levels of intolerance of uncertainty.*

2.3. Framing and the Construction of Subjective Value

Managerial judgments also depend on the reference point from which outcomes are evaluated. Prospect-based approaches challenge the assumption that decision makers evaluate equivalent gains and losses symmetrically. Instead, outcomes are judged relative to a reference point, and the psychological meaning of moving above that point can differ from the meaning of falling below it. Contemporary research continues to find robust framing effects in risky decisions, although their magnitude varies considerably across tasks and contexts (Balart et al., 2022; Ding et al., 2025).

Framing is particularly relevant to management because the same decision can often be communicated either as an opportunity to secure gains or as an attempt to prevent losses. A market-entry decision may be framed as the possibility of gaining a new revenue stream or as the risk of losing competitive position by remaining inactive. Likewise, a technology investment can be presented as an opportunity to improve productivity or as a means of avoiding future obsolescence. These frames need not change the underlying economic structure of the decision, yet they can alter how alternatives are evaluated.

Recent evidence shows that framing effects extend beyond simple lottery choices. Hsee et al. (2024) demonstrate that framing changes strategic interaction, while Treiman and Kool (2024) show that context preferences can shape subsequent decisions. Individual differences can also amplify or attenuate framing responses; even characteristics such as numeracy do not necessarily eliminate framing and may sometimes intensify it (Fitz et al., 2024). Confidence itself can also be frame-sensitive, suggesting that framing may influence not only what people choose but how certain they feel about their judgments (Sakamoto & Miyoshi, 2024).

For the present study, the important question is whether ambiguity carries the same subjective-value penalty in gain and loss domains. Existing theory does not justify assuming perfect symmetry. Indeed, the interaction between ambiguity and loss sensitivity may produce different patterns across contexts (Hu et al., 2023). We therefore expect the consequences of ambiguity for subjective value to depend on whether the managerial decision is represented as obtaining gains or avoiding losses.

H4. *Gain versus loss framing will moderate the relationship between ambiguity and subjective value, such that the effect of ambiguity on subjective value will differ across gain- and loss-framed managerial decisions.*

The hypothesis is deliberately framed as a difference rather than imposing a universal directional prediction. Recent work shows that framing effects are robust but context-sensitive (Ding et al., 2025), and the purpose of the present design is to identify how ambiguity and framing jointly shape managerial valuation rather than assume that loss framing must always strengthen or weaken ambiguity aversion.

2.4. Confidence, Calibration, and Managerial Decision Commitment

Subjective value alone does not determine whether a manager commits to a decision. Managers must also judge whether they trust their own evaluation. Confidence serves this metacognitive function by providing an internal assessment of decision reliability. Contemporary research treats confidence not merely as an emotion attached to choice but as information that can guide subsequent behavior, effort allocation, and task prioritization (Aguilar-Lleyda & de Gardelle, 2021; Balsdon & Philiastides, 2024; Fleming, 2024).

The difficulty is that confidence is noisy. Boundy-Singer et al. (2023) characterize confidence as an imperfect estimate of decision reliability, while Double and Birney (2024) show that making confidence judgments can itself influence decision processes. Non-diagnostic information may also alter confidence despite offering little objective basis for better decisions (Kohl et al., 2024). In managerial contexts, this distinction is critical because confidence often facilitates commitment, persuasion, and action even when it does not reflect superior judgment.

Management research provides substantial evidence that excessive confidence can have strategic consequences. Reviews of upper-echelons research document persistent interest in managerial overconfidence and its organizational implications (Brunzel, 2021). Overconfidence has been linked to organizational resilience, strategic risk taking, and performance, although its effects are not uniformly harmful (Burkhard et al., 2023; Kunz & Sonnenholzner, 2023). Some degree of confidence may support decisive action under uncertainty, while excessive confidence may encourage managers to underestimate downside risk or disregard contradictory evidence. Comparisons of entrepreneurs, managers, and employees likewise suggest that strategic decision makers differ systematically in optimism and overconfidence (Chochoiek et al., 2025).

Recent behavioral evidence provides an especially important insight for the present study: confidence may partly emerge *from the act of deciding*. Boyle et al. (2025) show that decision processes can generate overconfidence, and Boyle and Nye (2026) demonstrate that information distortion can contribute to overconfidence in managerial decisions. Thus, high confidence should not automatically be interpreted as evidence of good calibration. Managers may become more committed because they feel certain even when that certainty is not justified by their actual performance.

This matters for managerial decision commitment, defined here as the willingness to proceed with a selected option and allocate organizational resources to it. Subjective value should be one direct antecedent: an alternative that is perceived as more valuable should generate greater willingness to commit. Managerial experience and feedback histories may influence this relationship, but prior experience does not guarantee adaptive learning; managers can respond maladaptively even when they have accumulated substantial experience (Gaba et al., 2023). Similarly, executive confidence may sometimes inspire action but may also create organizational costs when certainty becomes excessive (Cohee & Barnhart, 2024; Schneck & Hautz, 2024).

H5. *Subjective value will be positively associated with managerial decision commitment.*

Because perceived ambiguity is expected to reduce subjective value, subjective value should also provide the mechanism through which perceived ambiguity ultimately affects commitment.

H6. *Perceived ambiguity will show a significant statistical indirect association with decision commitment through the highest pre-choice subjective-value rating across the two options* (BestPreChoiceValue).

Confidence should exert an additional influence. A manager who feels more certain about a judgment is generally more willing to act on it, even after accounting for the attractiveness of the underlying alternative.

H7. *Decision confidence will be positively associated with managerial decision commitment beyond the effect of subjective value.*

The more novel issue concerns whether confidence is *calibrated*. Confidence calibration refers to the correspondence between subjective confidence and observed decision performance. This distinction prevents high confidence from being treated automatically as high-quality judgment. Meta-analytic and strategic-management evidence suggests that overconfidence is associated with greater strategic risk taking (Burkhard et al., 2023), while recent managerial experiments indicate that distorted information processing can produce high confidence that is not fully warranted (Boyle & Nye, 2026). We therefore expect confidence to translate especially strongly into commitment when confidence exceeds what a manager's observed performance would justify.

H8. *Confidence calibration will moderate the relationship between confidence and managerial decision commitment, such that the positive relationship between confidence and commitment will be stronger when calibration is poorer, indicating overconfident commitment.*

This hypothesis captures a central managerial tension. Confidence can help managers act when information is incomplete, but poorly calibrated confidence may turn decisiveness into excessive commitment.

2.5. From Behavioral Theory to a Behaviorally Calibrated Fuzzy Subjective-Value Model

The preceding arguments create a methodological problem as well as a theoretical one. Conventional subjective-value models can account for nonlinear responses to gains, losses, and probabilities, but ambiguity introduces uncertainty about the probability information itself. Fuzzy approaches are well suited to representing imprecision, and a substantial literature has already integrated prospect-based valuation with different fuzzy-set structures (Tian et al., 2022). Recent work has developed models using interval uncertainty, spherical Z-numbers, regret-based fuzzy structures, probabilistic linguistic Bayesian networks, and related approaches (Ashraf et al., 2025; Pan et al., 2024; Wu et al., 2024; Y. Zhang et al., 2021). Consequently, the methodological contribution of the present study cannot rest simply on combining fuzzy sets with prospect theory.

A second development is equally important. Fuzzy decision research is moving away from treating linguistic membership functions as universally fixed. García-Zamora et al. (2024) introduced interpretable fuzzy-set construction through the Deck of Cards Method, while Dutta et al. (2026) extended this logic to higher-order uncertainty using a co-constructive approach. Consensus-based fuzzy linguistic scales have also been developed to accommodate differences across decision makers (García-Zamora et al., 2026). These advances show that personalization itself is no longer sufficient as a novelty claim.

At the same time, recent behavioral applications of fuzzy multi-criteria decision-making methods have become increasingly sophisticated. Yalvaç et al. (2026) use an advanced decision model to rank strategies for managing work–family conflict, while Eti et al. (2026) combine behavioral fuzzy logic with a decision framework for psychological determinants of consumer preferences. Such studies demonstrate that behavioral concepts can be embedded within fuzzy decision architectures. However, their dominant logic remains one in which experts or decision makers provide evaluations that are subsequently transformed, weighted, and ranked.

The present study takes a different route. Rather than asking experts to assign the importance of psychological factors, it uses repeated pre-choice ambiguity appraisals from an independent pilot as an empirical calibration target. The fuzzy hesitation/ambiguity-response term is calibrated against normalized perceived ambiguity using objective ambiguity, ambiguity width, intolerance of uncertainty, and the ambiguity × intolerance-of-uncertainty interaction. In this sense, the fuzzy representation is behaviorally calibrated because observed appraisal responses anchor the mapping rather than because participants select predefined linguistic membership functions. Confidence calibration remains a separate metacognitive construct used to test H8 and is not an input to the primary π equation.

This distinction is consistent with the broader subjective-value literature, which increasingly treats valuation as contextual and person-specific rather than as a fixed property of an option (Soukupova et al., 2026; Zheng et al., 2026). It also preserves the central confidence issue identified in metacognitive research: confidence contains useful information, but that information is noisy and must be interpreted relative to performance (Boundy-Singer et al., 2023; Fleming, 2024). In the present study, that confidence-calibration problem is examined behaviorally in H8 rather than used to fit the primary fuzzy hesitation/ambiguity-response mapping.

Accordingly, the proposed Behaviorally Calibrated Fuzzy Subjective-Value (BCFSV) model is built around three layers. First, objective decision information distinguishes measurable risk from ambiguity and, within ambiguous trials, captures the width of the probability interval. Second, an independently pilot-calibrated hesitation/ambiguity-response term varies with ambiguity exposure, ambiguity width, intolerance of uncertainty, and the ambiguity × intolerance-of-uncertainty interaction. Third, this π term attenuates the signed prospect-value difference and is evaluated against actual managerial choices and commitment. Confidence calibration is retained outside the primary π equation as a separate moderator of the confidence–commitment relationship. The model therefore connects behavioral appraisal with fuzzy representation without replacing observed behavior with expert-based rankings.

A useful test of this methodological contribution is not whether the proposed model merely generates different scores, but whether the nonlinear BCFSV transformation adds out-of-sample predictive information beyond the information from which it is constructed. Accordingly, the critical incremental test compares BCFSV with a same-information covariate benchmark that receives the corresponding raw inputs directly. Expected-value, conventional prospect-based, and objective-fuzzy models provide additional contextual benchmarks. This design separates the value of the BCFSV transformation from the value of simply adding more predictors.

H9. *Adding the Behaviorally Calibrated Fuzzy Subjective-Value transformation to a same-information covariate benchmark will provide incremental out-of-sample predictive value for managerial choice and decision commitment.*

2.6. Conceptual Model

Taken together, the hypotheses propose a sequential behavioral process. Objective ambiguity—including the presence of imprecise probability information and, within ambiguous trials, its degree—first influences how ambiguous the decision is perceived to be. This effect varies with intolerance of uncertainty. Perceived ambiguity then shapes pre-choice subjective value, while gain versus loss framing conditions valuation. The highest of the two pre-choice subjective-value ratings (BestPreChoiceValue) predicts decision commitment. Confidence provides an additional post-choice pathway to commitment, and its effect is conditioned by the participant-specific calibration marker.

The core theoretical pathway is therefore:

Objective Ambiguity → Perceived Ambiguity → Subjective Value → Decision Commitment. Intolerance of uncertainty moderates the first link, whereas gain/loss framing conditions the effect of ambiguity on subjective value. Confidence has a direct positive relationship with decision commitment, and confidence calibration moderates this relationship by identifying circumstances in which high confidence reflects overconfident rather than well-calibrated commitment.

The methodological layer is connected to this behavioral framework but is analytically distinct. Independent-pilot ambiguity appraisals are used to calibrate the individualized π component of BCFSV, while confidence calibration is examined separately in H8. The main experiment tests the behavioral pathways, and the predictive analyses then evaluate whether the pilot-calibrated BCFSV transformation adds information beyond exact same-information and richer behavioral benchmarks. Figure 1 presents the integrated behavioral model and the BCFSV methodological layer used in the analysis.

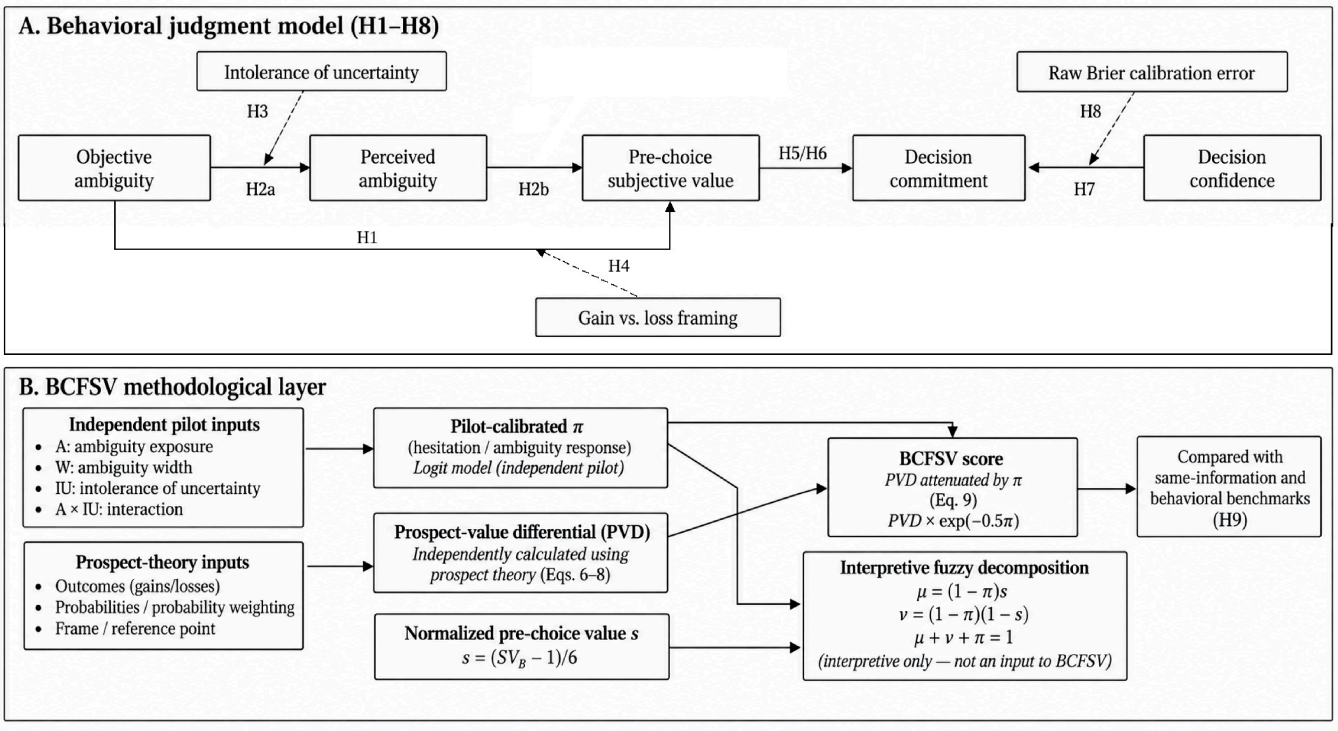


Figure 1. Integrated behavioral and BCFSV methodological model. (**A**) summarizes H1–H8, including the moderation roles of intolerance of uncertainty, gain/loss framing, and Raw Brier calibration error. Solid arrows represent hypothesized directional relationships between constructs, whereas dashed arrows indicate moderating effects (H3, H4, and H8). (**B**) shows the independent-pilot calibration of π and the independent calculation of the prospect-value differential (PVD), which jointly determine the BCFSV score through the attenuation specified in Equation (9). Solid arrows in panel B indicate

computational dependencies and model-evaluation steps. The μ/ν components form a separate interpretive fuzzy decomposition and do not enter the BCFSV score. The resulting BCFSV score is evaluated against the predictive benchmarks. Arrow styles distinguish the roles of the relationships rather than their statistical significance. In panel A, solid arrows represent hypothesized directional relationships between constructs, whereas dashed arrows indicate moderating effects (H3, H4, and H8). In panel B, solid arrows indicate computational dependencies and model-evaluation steps. Arrow styles distinguish the roles of the relationships rather than their statistical significance.

3. Materials and Methods

The study used a scenario-based behavioral experiment conducted in three stages: expert validation, pilot testing, and the main experiment. Seven experts evaluated the scenario variants, 42 working adults completed an independent pilot, and the main analytical sample comprised 525 organizational decision makers. The design combined repeated managerial choices with pre-choice appraisal measures, post-choice confidence and commitment, and a separate behavioral calibration block.

3.1. Research Design

A 2 × 2 repeated-measures factorial design manipulated uncertainty type (risk vs. ambiguity) and decision frame (gain vs. loss). Participants evaluated 12 managerial scenarios using one of four counterbalanced forms (A–D), with three trials in each condition. Scenario order was randomized within participant. Across the analytical sample, the design yielded 6300 trials, equally divided between risk and ambiguity and between gain and loss framing. Within each scenario, the gain- and loss-framed variants preserved the same probabilities and monetary quantities, so the framing contrast changed the reference description rather than the underlying numerical structure. Expert validation was conducted from 18 April to 20 April 2026, with seven experts evaluating all 48 scenario variants. Framing clarity was rated highly (overall $M = 4.43$ on a five-point scale; variant-level means ranged from 3.86 to 4.71). An independent pilot with 42 working adults was conducted from 22 April to 27 April 2026, yielding 504 pilot trial observations. These stages provided design-level and pilot evidence for the scenario materials; participant-level framing-recognition accuracy is not claimed because corresponding recognition responses are not available in the retained main-study analysis dataset. Table 1 summarizes the experimental design, validation stages, pilot evidence, and analytical sample.

Table 1. Experimental design and validation.

Component	Specification
Factorial design	2 (risk vs. ambiguity) × 2 (gain vs. loss), repeated measures
Managerial scenarios/trials per participant	12/12
Counterbalanced forms	4 (A–D); 3 trials per experimental condition
Expert validation	7 experts; 18–20 April 2026; S-CVI/Ave = 0.955
Pilot sample	42 participants; 22–27 April 2026; 504 trials; independent BCFSV calibration sample
Pilot perceived-ambiguity reliability	Spearman–Brown = 0.769
Pilot decision-commitment reliability	Spearman–Brown = 0.821
Pilot ambiguity manipulation	Risk $M = 2.97$; ambiguity $M = 3.83$; $d = 1.13$
Main analytical sample/trials	525 participants/6300 trials
Framing design validation	12/12 scenario pairs numerically matched; 7-expert clarity $M = 4.43/5$

Because the same scenarios were evaluated across participants, the data were treated as cross-classified by participant and scenario rather than as independent trial observations.

3.2. Participants and Procedure

Data collection for the main study was conducted in Türkiye from 5 May to 30 May 2026, using an online questionnaire administered through Google Forms. Recruitment targeted employed adults with organizational decision-making responsibilities. Invitations were distributed through company email networks, professional mailing lists, social media accounts, and professional networking platforms; each invitation described the project as a study of managerial judgment and organizational decision-making and provided a direct questionnaire link. Participation was voluntary and uncompensated. Eligibility required current professional activity together with responsibility for managerial, supervisory, project, budgetary, resource-allocation, or comparable organizational decisions, and these criteria were screened before entry into the experimental tasks.

After consent and eligibility screening, participants were assigned to one of four counterbalanced forms and completed 12 randomized managerial scenarios. Within each trial, perceived ambiguity was assessed first, followed by pre-choice attractiveness ratings for Options A and B, the behavioral choice, post-choice decision confidence, and decision commitment. After all 12 managerial trials, participants completed 12 objectively scored confidence-calibration decisions and three objective numeracy checks, followed by the intolerance-of-uncertainty and general risk-propensity measures and then demographic and professional-background questions. The complete wording, response formats, correct answers, and scoring rules for the three numeracy checks are provided in Table S3 of the Supporting Information. Because intolerance of uncertainty and the calibration/numeracy measures were collected after the managerial scenarios, analyses using those participant-level variables are interpreted as associational or as held-out-participant generalization rather than as prospective decision-time prediction. Among the 525 retained respondents, completion time averaged 19.08 min ($SD = 4.08$; median = 19.28; range = 7.00–32.61 min).

Data-quality screening was applied sequentially at the participant level: eligibility and organizational decision responsibility, survey completion, attention-check performance, duplicate/duplicate-like status, and response time. The archived screening flags identify 10 ineligible records, eight incomplete surveys, seven attention-check failures, four duplicate or duplicate-like responses, and six otherwise eligible completed responses excluded by the response-time screen. The six time-screen exclusions lasted 4.04–6.25 min, whereas the shortest retained completion was 7.00 min, supporting an operational threshold of less than 7 min. The archived data preserve the attention pass/fail and duplicate flags; the exact participant-facing attention-check wording and a separate duplicate-matching algorithm specification were not retained and are not reconstructed. In total, 35 of 560 screened records were excluded, leaving 525 participants (93.8% retention).

No a priori power analysis was used to determine the recruitment target; the analytical sample was fixed after application of the documented screening rules. As a transparent participant-level sensitivity reference, a conventional two-sided test with $N = 525$ and $\alpha = 0.05$ has 80% power to detect an association of approximately $|r| = 0.12$. This reference is not treated as a multilevel power calculation because sensitivity for repeated-measures interactions depends on clustering and the random-effects structure. Participants had a mean age of 41.49 years ($SD = 9.28$), an average of 18.11 years of work experience ($SD = 10.29$), and 8.02 years of managerial or formal decision-making experience ($SD = 6.39$). Formal managers represented 476 participants (90.7%); the remaining participants were professionals with documented organizational decision responsibility. The largest industry groups were logistics and transportation (12.8%), professional services (12.6%), technology (11.4%),

finance (11.2%), and manufacturing (10.3%). Table 2 summarizes the demographic and professional characteristics of the analytical sample. Throughout the manuscript, N denotes a complete sample or dataset count, whereas n denotes the count within a subgroup, category, or experimental condition.

Table 2. Main-sample characteristics.

Characteristic	Count/Mean	%/*SD*
Analytical sample	525	100%
Age (years)	41.49	$SD = 9.28$
Work experience (years)	18.11	$SD = 10.29$
Managerial/decision experience (years)	8.02	$SD = 6.39$
Formal managers	476	90.7%
Women	258	49.1%
Men	252	48.0%
Non-binary/another identity	9	1.7%
Prefer not to say	6	1.1%

3.3. Experimental Scenarios

The 12 scenarios covered market entry, technology upgrading, product launch, supplier partnership, R&D, capacity expansion, pricing, sustainability investment, digital transformation, strategic alliances, talent investment, and cybersecurity. Option A provided the more certain outcome, whereas Option B offered a potentially different expected value under either a precise probability (risk) or a probability interval (ambiguity). Ambiguity intervals had widths of 0.20, 0.30, or 0.40 around the same midpoint used in the corresponding risk condition. Recalculation from the archived scenario parameters showed an economically balanced bank: four scenarios favored Option B, four were approximately equivalent, and four favored Option A in expected-value terms. Approximate equivalence was defined as an absolute Option B expected-value advantage of less than $1, thereby treating only floating-point rounding differences as zero. Gain- and loss-framed versions retained the same probabilities and monetary quantities.

The within-trial measurement order followed the proposed process sequence and preserved temporal separation among the focal measures. Participants first evaluated perceived ambiguity and the attractiveness of both options before choosing; confidence and commitment were measured only after the choice. This ordering ensured that both subjective-value ratings were available independently of the later behavioral choice. Table 3 summarizes the sequence from perceived ambiguity and pre-choice valuation to choice, confidence, and commitment.

Table 3. Trial-level measurement sequence.

Order	Construct	Operationalization	Scale
1	Perceived ambiguity	Uncertainty and difficulty in evaluating Option B	2 items, 1–7
2	Subjective value	Pre-choice attractiveness of Options A and B	2 ratings, 1–7
3	Behavioral choice	Selected managerial option	Option A/Option B
4	Decision confidence	Confidence that the selected option is better	0–100
5	Decision commitment	Implementation commitment and resource willingness	2 items, 0–100

3.4. Measures

Subjective value was measured before choice by asking participants to rate the attractiveness of both Option A and Option B on seven-point scales (1 = not at all attractive, 7 = extremely attractive), following direct attractiveness approaches in risky-choice research (Olschewski et al., 2025). The pre-choice value of Option B was used as the focal valuation outcome for H1, H2, and H4. For the commitment pathway, the primary pre-choice value construct was BestPreChoiceValue, defined as $\max(SV_A, SV_B)$. Because this quantity is determined entirely from the two ratings obtained before choice, it does not depend on which option was subsequently selected. The selected-option value variable was retained only for sensitivity analysis.

Perceived ambiguity was assessed before valuation using two seven-point items concerning uncertainty in the probability information and difficulty evaluating Option B, adapted from published ambiguity assessments (Toribio-Flórez et al., 2023). The items were averaged; main-sample Spearman–Brown reliability was 0.795. Decision confidence was recorded after choice on a 0–100 scale and followed established decision-confidence approaches (Arts et al., 2024; Boundy-Singer et al., 2023). Decision commitment was measured with two 0–100 items assessing implementation commitment and willingness to allocate organizational resources; their mean composite had a Spearman–Brown coefficient of 0.791 (Hur et al., 2026).

Intolerance of uncertainty was measured at the participant level using four items rated from 1 (strongly disagree) to 6 (strongly agree), based on the established IUS framework (Carleton et al., 2007; Valencia et al., 2025). Internal consistency was $\alpha = 0.786$. A one-factor maximum-likelihood measurement model produced standardized loadings from 0.604 to 0.730, composite reliability (CR) = 0.788, average variance extracted (AVE) = 0.482, $\chi^2(2) = 0.225$, $p = 0.894$, CFI = 1.000, RMSEA = 0.000, and SRMR = 0.004. The AVE value is therefore reported descriptively rather than used to claim strong convergent validity. General risk propensity was assessed with the validated 0–10 willingness-to-take-risks item (Dohmen et al., 2011). Managerial experience was the primary demographic control; age was reserved for robustness analyses because of its overlap with work experience.

Confidence calibration was estimated from 12 objectively scored managerial expected-value decisions comprising four easy, four medium, and four difficult tasks, with six decisions for which Option A was objectively correct and six for which Option B was objectively correct. Each decision was followed by a confidence judgment from 50% to 100%. Participant-level calibration was summarized using accuracy, raw Brier error, and overconfidence. Raw Brier error, standardized across participants, was the primary calibration moderator for H8 because it is directly derived from the objective task responses and confidence judgments. Three objective numeracy items were administered after the calibration decisions and scored 0/1; the summed NumeracyScore ranged from 0 to 3 ($M = 1.89$, $SD = 0.94$), with correct-response rates of 71.4%, 63.4%, and 53.7% for the first, second, and third numeracy checks, respectively. The complete administered wording and scoring information appear in Table S3. A numeracy-adjusted Brier residual was retained as a sensitivity measure rather than as the primary calibration indicator.

The calibration block was designed as a heterogeneous criterion-task battery rather than as a reflective scale: the 12 decisions intentionally varied in difficulty and were used to estimate how closely stated confidence tracked objective performance. Odd-even Spearman–Brown coefficients were 0.594 for accuracy, 0.422 for raw Brier error, 0.385 for numeracy-adjusted Brier error, and 0.371 for overconfidence. These coefficients are interpreted as indices of person-level score precision rather than evidence that the tasks form a unidimensional psychometric scale, and calibration-related effects are therefore interpreted conservatively. Robustness analyses re-estimated H8 using numeracy-adjusted Brier error,

overconfidence, and accuracy separately; these alternatives arise from the same 12-task battery and are treated as scoring-rule sensitivity checks rather than independent replications. Table 4 summarizes the study measures and their analytical roles.

Table 4. Summary of study measures.

Variable	Level	Measurement	Primary Role
Subjective value	Trial	Pre-choice attractiveness of A and B, 1–7; BestPreChoiceValue = max(A, B)	H1/H2/H4 outcome; H5/H6 predictor/mediator
Perceived ambiguity	Trial	2 items, 1–7	Mediator
Decision confidence	Trial	1 item, 0–100	Predictor
Decision commitment	Trial	2 items, 0–100	Outcome
Intolerance of uncertainty	Participant	4 items, 1–6	H3 moderator/BCFSV input
Raw Brier error	Participant	12 objective decisions + confidence; standardized raw Brier primary	H8 moderator
Risk propensity	Participant	1 item, 0–10	Primary control
Managerial experience	Participant	Years	Primary control
Numeracy score	Participant	3 objective items; 0–3 total	Ancillary adjustment/sensitivity

3.5. Behaviorally Calibrated Fuzzy Subjective-Value Model

The BCFSV framework represents the focal uncertain option using an intuitionistic fuzzy structure. For participant i in trial t, support, opposition, and hesitation/ambiguity response are denoted by μ, ν, and π, respectively, subject to the standard admissibility condition:

$$\tilde{A}_{it} = (\mu_{it}, \nu_{it}, \pi_{it}), \qquad \mu_{it} + \nu_{it} + \pi_{it} = 1 \tag{1}$$

Pre-choice subjective value of Option B was normalized to $s_{it} = (SV_{\mathrm{B},it} - 1)/6$. Given π_{it}, support and opposition were obtained algebraically without additional fitted parameters. These μ and ν components provide an interpretable decomposition of the focal option but are not treated as independently estimated predictive constructs:

$$s_{it} = \frac{SV_{\mathrm{B},it} - 1}{6}, \qquad \mu_{it} = (1 - \pi^{*}_{it})s_{it}, \qquad \nu_{it} = (1 - \pi^{*}_{it})(1 - s_{it}) \tag{2}$$

The primary hesitation mapping was calibrated using only the independent pilot sample (42 participants; 504 trials) before application to the main sample. The observed calibration criterion was perceived ambiguity normalized to the unit interval as (PerceivedAmbiguity − 1)/6. A fractional-logit generalized linear model with participant-cluster-robust standard errors related this criterion to ambiguity exposure A_t, centered ambiguity width W_t^{c} (−1 = narrow, 0 = medium, +1 = wide; risk = 0), pilot-standardized intolerance of uncertainty IU_{zi}, and $A_t \times IU_{zi}$. Intolerance of uncertainty was standardized using the pilot mean (3.690) and sample SD (0.660), and the same scaling was transported unchanged to the main sample. Pilot-only model comparison selected the $A + W + IU + A \times IU$ specification (AIC = 456.701); adding confidence-calibration error did not improve the pilot criterion model (AIC = 458.538) and was therefore excluded from the primary π mapping. The resulting coefficients were $\beta_0 = -0.721531$, $\beta_1 = 0.605064$, $\beta_2 = 0.222976$, $\beta_3 = 0.163682$, and $\beta_4 = 0.080596$. The pilot specification and β coefficients were established on 1 May 2026, before the main study began on 5 May 2026, and were subsequently applied unchanged

to the main sample; main-sample choice and commitment outcomes were not used for coefficient estimation, specification selection, or tuning. Thus, π is a model-derived fuzzy representation empirically anchored to observed pre-choice ambiguity appraisal rather than a separately administered hesitation scale.

$$\text{logit}(\pi_{it}) = \beta_0 + \beta_1 A_t + \beta_2 W_t^c + \beta_3 IU_{zi} + \beta_4 (A_t \times IU_{zi}) \quad (3)$$

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (4)$$

Confidence calibration was kept analytically separate from the primary hesitation equation. For H8, raw Brier error was defined from the 12 calibration decisions, where C_{ij} is reported confidence expressed as a probability and Y_{ij} equals 1 for a correct response and 0 otherwise. The participant-level raw Brier score was standardized for the primary moderation analysis; the numeracy-adjusted residual, overconfidence, and accuracy were used only in sensitivity analyses:

$$\text{RawBrier}_i = \frac{1}{12} \sum_{j=1}^{12} \left(C_{ij} - Y_{ij} \right)^2 \quad (5)$$

For risk trials, probability weighting used the stated probability. For ambiguity trials, the midpoint of the probability interval entered the weighting function, while interval width was represented separately in π:

$$p_t = \begin{cases} p_t^{\text{known}}, & \text{risk}, \\ \frac{p_{\text{low},t} + p_{\text{high},t}}{2}, & \text{ambiguity}, \end{cases} \quad (6)$$

Reference-dependent outcome valuation and probability weighting followed the functions below. Monetary outcomes were first expressed in units of \$100,000. The verified prospect parameters were $\alpha = 0.88$, $\beta = 0.88$, $\lambda = 1.85$, and $\gamma = 0.68$. In gain-framed scenarios, the reference point was the current plan/status quo: Option A represented a certain improvement and Option B a larger improvement with probability p and zero improvement otherwise. In loss-framed scenarios, the reference point was the organizational target (zero remaining shortfall): Option A was coded as the remaining loss $-(B - A)$, whereas Option B produced zero remaining loss with probability p and the full loss $-B$ otherwise. The full-loss outcome was weighted by $1 - w(p)$, not by $w(1 - p)$. Using these rules reproduced all 6300 archived ProspectValueDifference values to six-decimal rounding (maximum absolute reconstruction error = 4.97×10^{-7}).

$$v(x) = \begin{cases} x^{\alpha}, & x \geq 0, \\ -\lambda(-x)^{\beta}, & x < 0, \end{cases} \quad (7)$$

$$w(p) = \frac{p^{\gamma}}{\left[p^{\gamma} + (1 - p)^{\gamma} \right]^{1/\gamma}} \quad (8)$$

The signed prospect-value difference between Options B and A was then attenuated by the pilot-calibrated π term:

$$\text{BCFSV}_{it} = \Delta PV_{it} \, e^{-\theta \pi_{it}^*}, \qquad \Delta PV_{it} = V_{\text{B},it}^{\text{PT}} - V_{\text{A},it}^{\text{PT}} \quad (9)$$

The μ and ν components are retained for interpretive decomposition, whereas the BCFSV score uses π to attenuate the absolute magnitude of the signed prospect-value differential toward zero. The attenuation scale was fixed at $\theta = 0.500$ and was not optimized against choice or commitment outcomes. Sensitivity analyses using $\theta = 0.25$, 0.75, and 1.00 produced BCFSV scores correlated at least 0.995 with the $\theta = 0.50$ specification, indicating that the score ordering was not narrowly dependent on the selected attenuation scale. Pilot

calibration showed $r = 0.632$, RMSE = 0.113, and MAE = 0.090; participant-grouped and leave-one-scenario-out pilot validation produced mean correlations of 0.616 and 0.612, respectively. Applying the fixed pilot coefficients to the 6300 main-study trials yielded external criterion correspondence of $r = 0.620$, RMSE = 0.118, and MAE = 0.094. Predictive evaluation of the resulting BCFSV score is therefore treated separately from validation of the π mapping itself.

3.6. Data Analysis

Analyses proceeded from screening and measurement checks to hypothesis testing, BCFSV validation, and predictive comparison. Screening followed the documented eligibility, completion, attention-check, duplicate, and response-time rules in Section 3.2 and Tables S8, S20 and S22. Descriptive statistics, reliability estimates, manipulation checks, and the IUS-4 one-factor measurement model were examined before hypothesis testing. Primary inferential models treated the 12 validated scenarios as fixed experimental stimuli through 11 scenario indicator terms and modeled repeated observations through participant random effects. H1 and H4 used participant random intercepts plus ambiguity slopes; H2 path *a* used the same structure; H2 path *b* and H3 used participant random intercepts because the attempted mediator/ambiguity slope fits were singular. H5 and H6 used participant random intercepts plus the focal pre-choice-value slope, and H7 and H8 used participant random intercepts plus the confidence slope. A generic continuous-outcome specification is:

$$Y_{is} = \beta_0 + X_{is}\beta + u_i + \gamma_s + \varepsilon_{is} \quad (10)$$

For the predictive validation of binary managerial choice, models were estimated within participant-grouped training folds using logistic regression of the form:

$$\text{logit}\{\text{P}(\text{ChoiceB}_{is} = 1)\} = \beta_0 + X_{is}\beta \quad (11)$$

Decision commitment was evaluated with corresponding linear prediction models within the same grouped folds:

$$\text{Commitment}_{is} = \beta_0 + X_{is}\beta + \varepsilon_{is} \quad (12)$$

H2 and H6 were evaluated as product-of-coefficients statistical indirect effects. H2 tested ambiguity → perceived ambiguity → SubjectiveValue_B_Pre. H6 tested perceived ambiguity → BestPreChoiceValue → commitment, where BestPreChoiceValue is defined before choice and therefore avoids conditioning the mediator on the subsequently selected option. For both pathways, the indirect effect $a \times b$ and its uncertainty were estimated with 5000 participant-cluster bootstrap resamples; all 12 trials for a sampled participant were resampled together, both path models were re-estimated within each bootstrap replicate, and scenario fixed effects were retained. This procedure preserves the sampling covariance between the *a* and *b* paths. Because the mediators themselves were not randomized, both H2 and H6 are interpreted as statistical indirect effects rather than identified causal mediation. Moderation was evaluated through ambiguity × intolerance of uncertainty (H3), ambiguity × assigned frame (H4), and confidence × standardized raw Brier error (H8). Risk propensity and managerial experience were the primary controls in commitment models; age was added only in robustness analysis.

Predictive validation was divided into two analytically distinct families to respect variable timing. The strict pre-choice family used only information available by the focal choice: expected value, prospect value, an objective non-personalized fuzzy score, and a pre-choice behavioral benchmark containing expected/prospect information, ambiguity and width, assigned frame, perceived ambiguity, and the two pre-choice subjective-value ratings. The second family evaluated held-out-participant generalization for BCFSV and related

benchmarks; these models could use intolerance of uncertainty and other participant-level information collected after the 12 managerial scenarios and therefore are not interpreted as prospective decision-time prediction. The exact same-information benchmark entered the raw BCFSV inputs directly (prospect value, ambiguity, width, pilot-scaled IU, and ambiguity × IU), and the incremental test added the BCFSV transformation to those same predictors. An augmented benchmark additionally incorporated the relevant behavioral/person-level information used in the broader comparison. Five-fold GroupKFold cross-validation split observations by participant so that all 12 trials from a participant appeared exclusively in training or test data. All scaling, preprocessing, and model coefficients were estimated within training folds only. Choice performance was evaluated with AUC, log loss, and Brier score; commitment performance used MAE, RMSE, and out-of-sample R^2. Incremental differences between matched models were summarized with 2000 participant-cluster bootstrap resamples of paired out-of-fold predictions.

Robustness analyses examined whether the substantive conclusions changed under alternative calibration indicators (numeracy-adjusted Brier error, overconfidence, and accuracy), a stricter response-time screen excluding the fastest approximately 5% of retained participants, gain- and loss-framed subsamples, a formal-manager-only sample, lower versus higher managerial-experience groups, age adjustment, alternative covariance treatment, alternative hesitation-attenuation functions, and θ values from 0.25 to 1.00. These analyses were treated as sensitivity checks rather than opportunities to redefine the primary specification after observing significance. All statistical tests were two-tailed with $\alpha = 0.05$.

Software and computational environment. The reproducibility environment used stable software releases specified for 4 September 2026. IBM SPSS Statistics 32.0.0 was used for screening verification, descriptive summaries, conventional reliability statistics, and data-quality checks. Inferential mixed-effects, interaction, marginal-effect, bootstrap, and measurement-model analyses were implemented in R 4.6.1 using lme4 2.0-6, lmerTest 3.2-1, emmeans 2.0.4, boot 1.3-32, and performance 0.18.1. Data restructuring, calibration-score construction, BCFSV computation, participant-grouped cross-validation, predictive-model comparison, and figure generation were implemented in Python 3.14.7 using NumPy 2.5.2, pandas 3.0.5, SciPy 1.18.1, scikit-learn 1.9.0, statsmodels 0.15.0, and Matplotlib 3.11.1. Tables S6 and S20–S22 of the Supporting Information provide numerical summaries, methodological documentation, the results summary, and the analysis specification.

4. Results

4.1. Sample Characteristics and Preliminary Analyses

A total of 560 individuals entered the main study. Thirty-five cases (6.3%) were excluded according to the documented screening criteria: 10 did not meet the organizational decision-responsibility criterion, eight did not complete the survey, seven failed the attention check, four were identified as duplicate or duplicate-like responses, and six fell below the response-time screen. The analytical sample therefore consisted of 525 participants and 6300 trial-level managerial decisions.

Participants had a mean age of 41.49 years ($SD = 9.28$), an average of 18.11 years of work experience ($SD = 10.29$), and 8.02 years of managerial or formal decision-making experience ($SD = 6.39$). Women represented 49.1% of the sample and men 48.0%; 1.7% identified as non-binary/another identity and 1.1% preferred not to report gender. Most participants held formal managerial or leadership roles (90.7%). The sample also covered a broad range of sectors, with the largest groups drawn from logistics and transportation, professional services, technology, finance, and manufacturing.

Descriptive statistics and measurement evidence are reported in Table 5. The four-item intolerance-of-uncertainty measure showed acceptable internal consistency ($\alpha = 0.786$). A

one-factor maximum-likelihood CFA yielded standardized loadings from 0.604 to 0.730, composite reliability of 0.788, and average variance extracted of 0.482; model fit was strong, $\chi^2(2) = 0.225$, $p = 0.894$, CFI = 1.000, RMSEA = 0.000, and SRMR = 0.004. Because AVE was slightly below 0.50, it is reported descriptively rather than used as evidence of strong convergent validity. The two-item perceived-ambiguity and decision-commitment measures showed Spearman–Brown coefficients of 0.795 and 0.791, respectively.

Table 5. Descriptive statistics and reliability estimates.

Variable	*M*	*SD*	Range	Reliability
Intolerance of uncertainty	3.49	0.77	1.25–6.00	$\alpha = 0.786$
Risk propensity	5.04	1.84	0–10	—
Perceived ambiguity	3.41	0.89	1.00–6.50	SB = 0.795
Subjective value of Option B	4.63	0.98	1–7	—
Decision confidence	72.41	10.21	35.88–99.00	—
Decision commitment	62.08	11.25	20.50–100.00	SB = 0.791
Option B choice	0.45	0.50	0–1	—

Note. Trial-level descriptive statistics are based on 6300 observations. SB = Spearman–Brown coefficient. IUS-4 CFA statistics are reported in the text and Supporting Information.

The confidence-calibration battery showed the intended difficulty gradient. Mean accuracy declined from 0.713 for easy tasks to 0.620 for medium tasks and 0.555 for hard tasks, while mean Brier error increased from 0.201 to 0.246 and 0.270. The odd-even Spearman–Brown coefficients were 0.594 for accuracy, 0.422 for raw Brier error, 0.385 for numeracy-adjusted Brier error, and 0.371 for overconfidence. These values indicate modest person-level score precision. Accordingly, Raw Brier error is treated as the primary calibration metric for H8 because it is directly task-derived and showed greater split-half precision than the adjusted Brier marker; the remaining calibration scores are used as robustness specifications rather than as independent replications.

Descriptive Pearson correlations showed that perceived ambiguity was negatively associated with the pre-choice subjective value of Option B ($r = -0.430$, $p < 0.001$), decision confidence ($r = -0.256$, $p < 0.001$), and decision commitment ($r = -0.240$, $p < 0.001$). Option B subjective value ($r = 0.327$, $p < 0.001$), the highest pre-choice subjective value across the two options ($r = 0.472$, $p < 0.001$), and decision confidence ($r = 0.326$, $p < 0.001$) were positively associated with commitment. At the participant level, intolerance of uncertainty correlated positively with mean perceived ambiguity ($r = 0.709$, $p < 0.001$) and negatively with mean Option B subjective value ($r = -0.416$, $p < 0.001$). The first six correlations use 6300 repeated trial observations and the last two use 525 participant-level observations. Trial-level Pearson *p*-values are unadjusted for within-participant dependence and are reported descriptively; the multilevel models below provide the primary inferential tests.

Overall, the focal behavioral measures displayed adequate variation across participants and trials. The verified sample structure, balanced experimental cells, measurement precision checks, and descriptive distributions supported proceeding to the manipulation, hypothesis, and model-validation analyses.

4.2. Manipulation and Design Checks

Participants reported substantially higher perceived ambiguity in ambiguous trials ($M = 3.85$, $SD = 0.82$) than in risk trials ($M = 2.96$, $SD = 0.73$), corresponding to a large

standardized difference ($d = 1.14$). The risk–ambiguity difference was statistically significant in the mixed-effects analysis ($p < 0.001$).

A mixed-effects model using the specified participant/scenario structure yielded a comparable ambiguity effect on perceived ambiguity ($b = 0.888$, $SE = 0.018$, 95% CI [0.852, 0.924], $p < 0.001$). Within ambiguous trials, wider probability intervals were also associated with greater perceived ambiguity ($b = 0.147$, $SE = 0.020$, $p < 0.001$). These results confirm sensitivity to both the presence and degree of ambiguity.

The framing factor was evaluated as a design-level check rather than as a participant-recognition manipulation check. Across all 12 scenarios, gain- and loss-framed versions retained identical probabilities and monetary quantities, with only the reference description changing. Expert-rated framing clarity was high (overall $M = 4.43/5$; variant-level means = 3.86–4.71). Experimental allocation was exactly balanced, with 1575 observations in each risk–gain, risk–loss, ambiguity–gain, and ambiguity–loss cell. The scenario bank was also balanced in expected-value terms: four scenarios favored Option B, four were approximately equivalent, and four favored Option A. Because participant-level frame-recognition responses were not retained, H4 is treated as the test of the assigned framing interaction rather than as evidence of recognition accuracy. Table 6 summarizes these manipulation and design checks.

Table 6. Ambiguity manipulation and framing-design checks.

Check	Result
Perceived ambiguity—Risk	$M = 2.96$, $SD = 0.73$
Perceived ambiguity—Ambiguity	$M = 3.85$, $SD = 0.82$
Ambiguity effect	$b = 0.888$, $SE = 0.018$, 95% CI [0.852, 0.924], $p < 0.001$
Cohen's d	1.14
Ambiguity-width effect	$b = 0.147$, $SE = 0.020$, $p < 0.001$
Expert-rated framing clarity	$M = 4.43/5$; variant means 3.86–4.71
Trials per experimental condition	$n = 1575$
Scenario expected-value balance	4 favor B/4 approximately equivalent/4 favor A
Framing numerical-equivalence check	12/12 scenario pairs matched on probabilities and monetary quantities
Participant framing-recognition accuracy	Not available in retained main-study data; not used as manipulation evidence

4.3. Hypothesis Testing

Hypotheses H1–H8 were estimated using the mixed-effects specifications documented in Section 3.6. Participant random slopes were retained when the models converged without singularity; where a planned slope was singular, the model used the documented simpler participant random-effects structure. The 12 validated scenarios were represented by fixed effects in the primary inferential models. H2 and H6 uncertainty was evaluated with 5000 participant-cluster bootstrap resamples in which both component paths were re-estimated jointly.

To facilitate interpretation of the effect sizes, perceived ambiguity and pre-choice subjective value were assessed on 1–7 scales, while decision confidence and decision commitment were assessed on 0–100 scales. A one-point difference in BestPreChoiceValue therefore refers to one attractiveness-scale point, whereas confidence effects are expressed per 10-point difference on the confidence scale and in points on the commitment scale. Raw Brier error was standardized for the moderation analysis; higher values indicate poorer confidence calibration. The estimated associations should not be interpreted as causal effects.

H1 was supported. Ambiguity significantly reduced the pre-choice subjective value assigned to Option B ($b = -0.396$, $SE = 0.021$, 95% CI $[-0.438, -0.355]$, $p < 0.001$). The descriptive means were consistent with this effect: subjective value was lower under ambiguity ($M = 4.43$) than under measurable risk ($M = 4.82$).

H2 examined the statistical indirect pathway from objective ambiguity to subjective value through perceived ambiguity. Ambiguity increased perceived ambiguity ($a = 0.888$, $SE = 0.018$, 95% CI $[0.852, 0.924]$, $p < 0.001$), and perceived ambiguity was negatively associated with Option B subjective value while ambiguity remained in the model ($b = -0.437$, $SE = 0.014$, 95% CI $[-0.466, -0.409]$, $p < 0.001$). The participant-cluster bootstrap indirect effect was −0.405 (bootstrap $SE = 0.017$, 95% CI $[-0.439, -0.373]$); the interval excluded zero. The residual direct effect of ambiguity was small and nonsignificant ($b = -0.008$, $SE = 0.023$, 95% CI $[-0.053, 0.037]$, $p = 0.730$). H2 was therefore supported as a statistical indirect effect, not as identified causal mediation.

H3 was also supported. The ambiguity × intolerance-of-uncertainty interaction predicted perceived ambiguity ($b = 0.181$, $SE = 0.016$, 95% CI $[0.149, 0.213]$, $p < 0.001$). The estimated ambiguity effect was 0.707 at −1 *SD* of intolerance of uncertainty ($SE = 0.023$, 95% CI $[0.662, 0.752]$), 0.888 at the mean ($SE = 0.016$, 95% CI $[0.856, 0.920]$), and 1.069 at +1 *SD* ($SE = 0.023$, 95% CI $[1.024, 1.114]$); all three simple slopes were statistically clear.

H4 predicted that the ambiguity effect would differ across assigned gain and loss frames. The ambiguity × loss-frame interaction was significant ($b = 0.208$, $SE = 0.041$, 95% CI $[0.128, 0.289]$, $p < 0.001$). The ambiguity-related reduction in subjective value was larger in gain-framed trials ($b = -0.501$, $SE = 0.030$, 95% CI $[-0.559, -0.443]$) than in loss-framed trials ($b = -0.292$, $SE = 0.030$, 95% CI $[-0.350, -0.234]$); both simple slopes were significant. H4 was therefore supported.

For H5, the primary pre-choice value predictor was redefined as BestPreChoiceValue, the higher of the two attractiveness ratings obtained before the participant made a choice. This choice-independent pre-choice value was strongly associated with decision commitment ($b = 6.359$, $SE = 0.163$, 95% CI $[6.040, 6.679]$, $p < 0.001$). A one-point increase in the best pre-choice value corresponded to approximately a 6.36-point increase in commitment, supporting H5.

H6 was re-estimated using the same choice-independent BestPreChoiceValue construct. Perceived ambiguity was negatively associated with BestPreChoiceValue ($a = -0.252$, $SE = 0.014$, 95% CI $[-0.279, -0.225]$, $p < 0.001$), and BestPreChoiceValue was positively associated with commitment while perceived ambiguity remained in the model ($b = 6.115$, $SE = 0.168$, 95% CI $[5.787, 6.444]$, $p < 0.001$). The participant-cluster bootstrap indirect effect was −1.584 (bootstrap $SE = 0.100$, 95% CI $[-1.784, -1.391]$), and the residual association between perceived ambiguity and commitment remained negative ($b = -1.105$, $SE = 0.173$, 95% CI $[-1.443, -0.767]$, $p < 0.001$). H6 was supported as a statistical indirect effect without conditioning the primary mediator on the later choice.

H7 was supported. After adjustment for BestPreChoiceValue, the experimental factors, risk propensity, and managerial experience, a 10-point increase in decision confidence on its 0–100 scale was associated with a 2.34-point increase in decision commitment on its 0–100 scale ($b = 2.335$, $SE = 0.142$, 95% CI $[2.057, 2.614]$, $p < 0.001$).

H8 was tested with standardized raw Brier error as the primary calibration indicator. The confidence × Raw Brier interaction was positive ($b = 0.874$, $SE = 0.125$, 95% CI $[0.630, 1.119]$, $p < 0.001$), indicating that the confidence–commitment association was stronger among participants with poorer confidence calibration. The estimated confidence slope was 1.48 commitment points per 10 confidence points at −1 *SD* of Raw Brier error ($SE = 0.182$, 95% CI $[1.124, 1.837]$), 2.36 at the mean ($SE = 0.137$, 95% CI $[2.087, 2.623]$), and 3.23 at +1 *SD* ($SE = 0.188$, 95% CI $[2.861, 3.598]$); all slopes were significant. H8 was there-

fore supported, with the calibration-score precision limitation retained in interpretation. Table 7 summarizes the H1–H8 estimates and inference.

Table 7. Summary of hypothesis tests.

Hypothesis	Tested Relationship	*b*	*SE*	95% CI	Result
H1	Ambiguity → subjective value	−0.396	0.021	[−0.438, −0.355]	$p < 0.001$; Supported
H2	Indirect: ambiguity → PA → subjective value	−0.405	0.017	[−0.439, −0.373]	Bootstrap CI excludes 0; Supported
H3	Ambiguity × intolerance of uncertainty → perceived ambiguity	0.181	0.016	[0.149, 0.213]	$p < 0.001$; Supported
H4	Ambiguity × loss frame → subjective value	0.208	0.041	[0.128, 0.289]	$p < 0.001$; Supported
H5	Best pre-choice value → commitment	6.359	0.163	[6.040, 6.679]	$p < 0.001$; Supported
H6	Indirect: PA → best pre-choice value → commitment	−1.584	0.100	[−1.784, −1.391]	Bootstrap CI excludes 0; Supported
H7	Decision confidence → commitment	2.335	0.142	[2.057, 2.614]	$p < 0.001$; Supported
H8	Confidence × Raw Brier error → commitment	0.874	0.125	[0.630, 1.119]	$p < 0.001$; Supported

Note. H2 and H6 report participant-cluster bootstrap indirect effects based on 5000 resamples; both component paths were re-estimated within each replicate. These effects are statistical indirect effects rather than identified causal mediation effects. Confidence in H7 and H8 is expressed per 10-point increase. Raw Brier error is standardized for H8; higher values indicate poorer calibration.

Figure 2 visualizes the two moderation results: the ambiguity effect on perceived ambiguity across intolerance-of-uncertainty levels and the confidence effect on commitment across Raw Brier calibration levels.

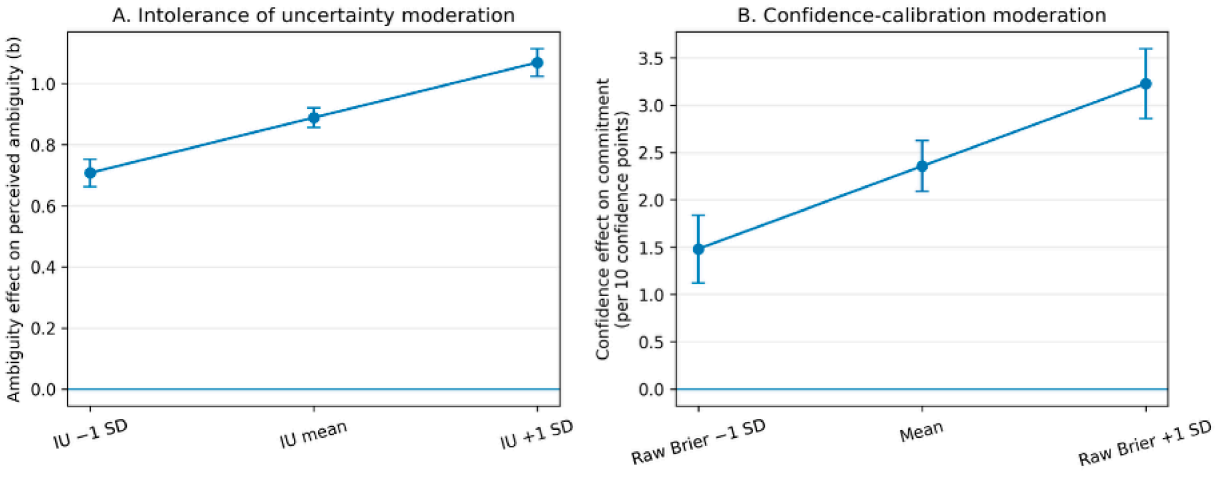


Figure 2. Moderation estimates. (**A**) shows the ambiguity effect on perceived ambiguity at −1 *SD*, the mean, and +1 *SD* of intolerance of uncertainty. (**B**) shows the confidence–commitment slope at −1 *SD*, the mean, and +1 *SD* of standardized Raw Brier error. Error bars show 95% confidence intervals.

4.4. Pilot Calibration and Validation of the BCFSV Model

The BCFSV hesitation component was calibrated empirically in the independent pilot sample ($N = 42$; 504 trials). The observed calibration criterion was normalized pre-choice

perceived ambiguity, defined as $(\text{PerceivedAmbiguity} - 1)/6$. A fractional-logit model was estimated using objective ambiguity, centered ambiguity width, pilot-standardized intolerance of uncertainty, and the ambiguity × intolerance-of-uncertainty interaction. No main-study choice or commitment outcome was used to estimate the calibration coefficients.

The pilot mapping was $\text{logit}(\pi) = -0.7215 + 0.6051A + 0.2230W + 0.1637IU + 0.0806(A \times IU)$. All five coefficients were statistically supported in the pilot using participant-cluster-robust inference. A calibration-error term was not included in the primary mapping: adding calibration error increased the pilot AIC from 456.70 to 458.54 and did not improve external criterion performance. Calibration error therefore remains a separate construct for H8 and sensitivity analyses rather than an input forced into the primary π equation.

Within the pilot sample, the pilot mapping showed moderate correspondence with observed normalized perceived ambiguity ($r = 0.632$, RMSE = 0.113, MAE = 0.090). Six participant-grouped held-out folds produced mean $r = 0.616$, RMSE = 0.114, and MAE = 0.091. Leave-one-scenario-out validation produced mean $r = 0.612$, RMSE = 0.114, and MAE = 0.092. These checks indicate that the mapping was not dependent on a small subset of pilot participants or scenarios.

The pilot coefficients were then transported to the main sample without refitting. Across the 6300 main-study trials, predicted π correlated 0.620 with observed normalized perceived ambiguity, with RMSE = 0.118 and MAE = 0.094. Predicted main-sample π values ranged from 0.210 to 0.723. Simulation of 2000 coefficient draws from the participant-cluster-robust pilot covariance matrix yielded a narrow external-correlation range (approximately 0.605–0.625 across the central 95% of draws), indicating that the external-validation result was stable to pilot coefficient uncertainty.

Using the pilot mapping, mean hesitation was 0.318 under risk and 0.454 under ambiguity. Mean support (μ) decreased from 0.436 under risk to 0.317 under ambiguity, whereas mean opposition (ν) was 0.246 and 0.229, respectively. Within ambiguous trials, mean π increased with interval width from 0.400 for narrow ambiguity to 0.453 for medium ambiguity and 0.508 for wide ambiguity. Because μ and ν are derived algebraically from π and normalized pre-choice subjective value, these components are interpreted as a fuzzy decomposition rather than as independently estimated latent variables. Table 8 summarizes the pilot coefficient estimates and validation statistics.

The attenuation scale θ remained fixed at 0.500 and was not optimized against the focal outcomes. Sensitivity specifications at $\theta = 0.25$, 0.75, and 1.00 produced BCFSV scores correlated at least 0.995 with the $\theta = 0.50$ score, and the incremental predictive conclusion reported in Section 4.5 was unchanged. Thus, the hesitation mapping is empirically anchored to pilot perceived-ambiguity responses while the attenuation strength remains a transparent fixed modeling choice.

The resulting BCFSV score preserves the sign of the prospect-value differential while shrinking its magnitude according to the pilot-calibrated hesitation representation. The role of π is therefore predictive/representational, whereas μ and ν provide an interpretable intuitionistic-fuzzy decomposition conditional on π and pre-choice subjective value. This distinction prevents the algebraically derived μ and ν terms from being presented as separate empirically validated predictors.

Overall, the calibration evidence supports the π mapping as a stable pilot-calibrated ambiguity-response representation. It does not, by itself, establish that the nonlinear BCFSV transformation improves prediction. Section 4.5 therefore evaluates that question directly against benchmarks containing the same underlying information.

Figure 3 summarizes the fuzzy decomposition and the increase in pilot-calibrated hesitation from risk to progressively wider ambiguity intervals.

Table 8. Independent-pilot BCFSV calibration and validation.

Component	Estimate/Performance	Validation/Interpretation
β_0 Intercept	-0.7215 ($SE = 0.0355$)	Pilot fractional-logit calibration
β_1 Ambiguity	0.6051 ($SE = 0.0346$)	Pilot; $p < 0.001$
β_2 Ambiguity width	0.2230 ($SE = 0.0384$)	Pilot; $p < 0.001$
β_3 IU (pilot standardized)	0.1637 ($SE = 0.0275$)	Pilot; $p < 0.001$
β_4 Ambiguity × IU	0.0806 ($SE = 0.0269$)	Pilot; $p = 0.0028$
Pilot apparent fit	$r = 0.632$; RMSE = 0.113; MAE = 0.090	504 pilot trials
Participant-grouped pilot validation	mean $r = 0.616$; RMSE = 0.114; MAE = 0.091	6 held-out participant folds
Scenario-held-out pilot validation	mean $r = 0.612$; RMSE = 0.114; MAE = 0.092	12 leave-one-scenario-out fits
Main-sample external validation	$r = 0.620$; RMSE = 0.118; MAE = 0.094	6300 main trials; pilot β fixed across samples

Note. The β coefficients were estimated in the independent pilot against normalized perceived ambiguity. IU was standardized using the pilot mean and sample *SD*. Calibration error is not part of the primary π mapping. Main-sample validation used the pilot coefficients without refitting against choice or commitment outcomes.

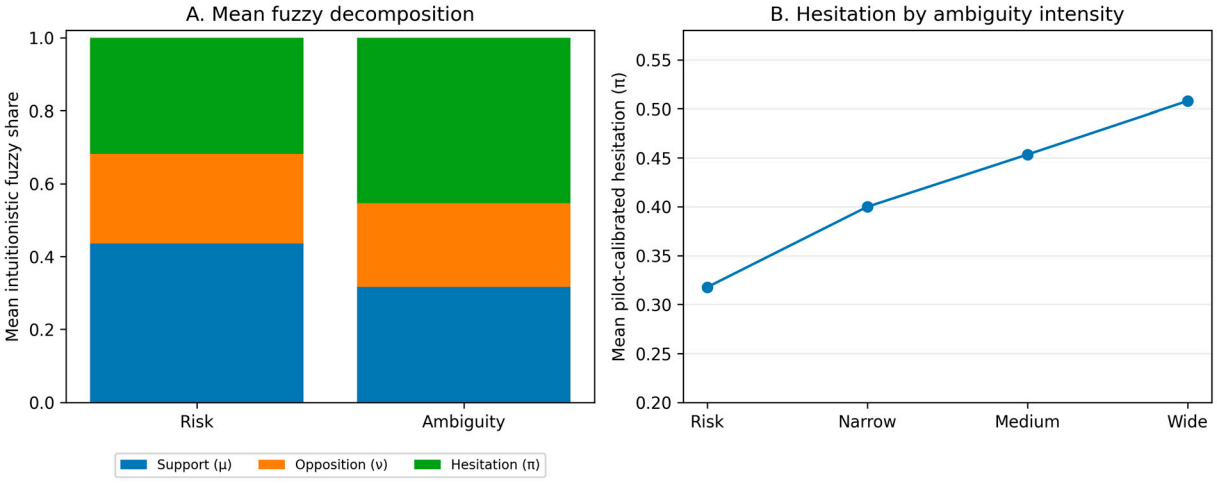


Figure 3. BCFSV fuzzy representation. (**A**) shows mean support (μ), opposition (ν), and hesitation (π) under risk and ambiguity using the pilot calibration. (**B**) shows mean π under risk and narrow, medium, and wide ambiguity.

4.5. Predictive Validation and H9

Predictive analyses were separated into two families to respect the measurement timeline. The first family evaluated strict pre-choice prediction using only information available by the focal choice: expected value, prospect value, an objective fuzzy score, and a pre-choice behavioral benchmark containing the experimental factors, perceived ambiguity, and the two pre-choice subjective-value ratings. The second family evaluated held-out-participant generalization using participant characteristics such as intolerance of uncertainty that were measured after the managerial trials. Accordingly, BCFSV analyses in the second family are not described as prospective decision-time prediction in the present experiment.

All predictive models used five-fold participant-grouped cross-validation, keeping each participant's 12 trials together. Preprocessing and model fitting were restricted to the training participants. Within the strict pre-choice family, the expected-value, prospect-

value, and objective-fuzzy models provided modest choice discrimination, whereas the pre-choice behavioral benchmark incorporated the appraisal and valuation information observed before the choice.

For choice, the mean fold AUC values were 0.600 for expected value, 0.611 for prospect value, 0.606 for the objective-fuzzy model, and 0.697 for the pre-choice behavioral benchmark. The corresponding mean log-loss values were 0.673, 0.670, 0.672, and 0.628, and mean Brier scores were 0.240, 0.239, 0.239, and 0.219. Thus, pre-choice behavioral appraisal and valuation information provided substantially better held-out choice discrimination than the conventional value-only representations.

For commitment, the strict pre-choice models yielded mean fold R^2 values of 0.002 for expected value, 0.003 for prospect value, 0.001 for the objective-fuzzy model, and 0.181 for the pre-choice behavioral benchmark. The pre-choice behavioral model also reduced MAE to 8.126 and RMSE to 10.169. These results show that information available before choice explained a meaningful, although incomplete, portion of later commitment variation.

The BCFSV score by itself did not show superior held-out-participant performance. Its mean choice AUC was 0.601 and its mean commitment R^2 was 0.001. The exact same-information raw benchmark, which entered the prospect-value difference, ambiguity, ambiguity width, pilot-scaled intolerance of uncertainty, and ambiguity × intolerance-of-uncertainty term directly, yielded choice AUC = 0.638 and commitment $R^2 = 0.034$. Adding the nonlinear BCFSV score produced choice AUC = 0.638 and commitment $R^2 = 0.036$.

Paired participant-bootstrap comparisons showed that these incremental differences were not distinguishable from zero. Relative to the same-information raw benchmark, adding BCFSV changed AUC by −0.00039 (95% CI [−0.00119, 0.00040]), log loss by +0.00023 (95% CI [−0.00018, 0.00065]), and Brier score by +0.00009 (95% CI [−0.00009, 0.00027]). For commitment, ΔMAE = −0.0074 (95% CI [−0.0179, 0.0028]), ΔRMSE = −0.0065 (95% CI [−0.0174, 0.0039]), and $\Delta R^2 = +0.00113$ (95% CI [−0.00069, 0.00302]). A richer augmented held-out benchmark produced the same conclusion: adding BCFSV changed choice AUC by −0.00009 and commitment R^2 by +0.00024, with both confidence intervals spanning zero. Table 9 reports the grouped cross-validation results.

Table 9. Participant-grouped predictive performance of the benchmark families.

Model	AUC	Log Loss	Brier Score	MAE	RMSE	R^2
1. Expected value (pre-choice)	0.600	0.673	0.240	8.902	11.230	0.002
2. Prospect value (pre-choice)	0.611	0.670	0.239	8.901	11.223	0.003
3. Objective fuzzy (pre-choice)	0.606	0.672	0.239	8.907	11.232	0.001
4. Pre-choice behavioral	0.697	0.628	0.219	8.126	10.169	0.181
5. BCFSV only	0.601	0.673	0.240	8.909	11.234	0.001
6. Same-information raw	0.638	0.660	0.234	8.757	11.043	0.034
7. Same-information + BCFSV	0.638	0.660	0.234	8.750	11.037	0.036
8. Augmented held-out benchmark	0.696	0.628	0.219	7.654	9.613	0.269
9. Augmented + BCFSV	0.696	0.628	0.219	7.653	9.612	0.269

Note. Values are mean five-fold participant-grouped cross-validation statistics. Models 1–4 use strict pre-choice information. Models 5–9 are held-out-participant generalization models; these may use participant characteristics measured after the 12 managerial trials and therefore are not interpreted as prospective focal-choice prediction. Lower log loss, Brier, MAE, and RMSE are better; higher AUC and R^2 are better.

H9 was not supported. The BCFSV transformation did not provide detectable incremental predictive information beyond a benchmark containing the same raw inputs, and the

conclusion was unchanged when BCFSV was added to a richer behavioral benchmark. The predictive contribution in this dataset is therefore attributable primarily to the underlying behavioral/appraisal information rather than to a demonstrable incremental advantage of the nonlinear BCFSV transformation itself.

Figure 4 displays the mean-fold choice AUC and commitment R^2 values, separating the strict pre-choice benchmarks from the held-out-participant BCFSV comparisons.

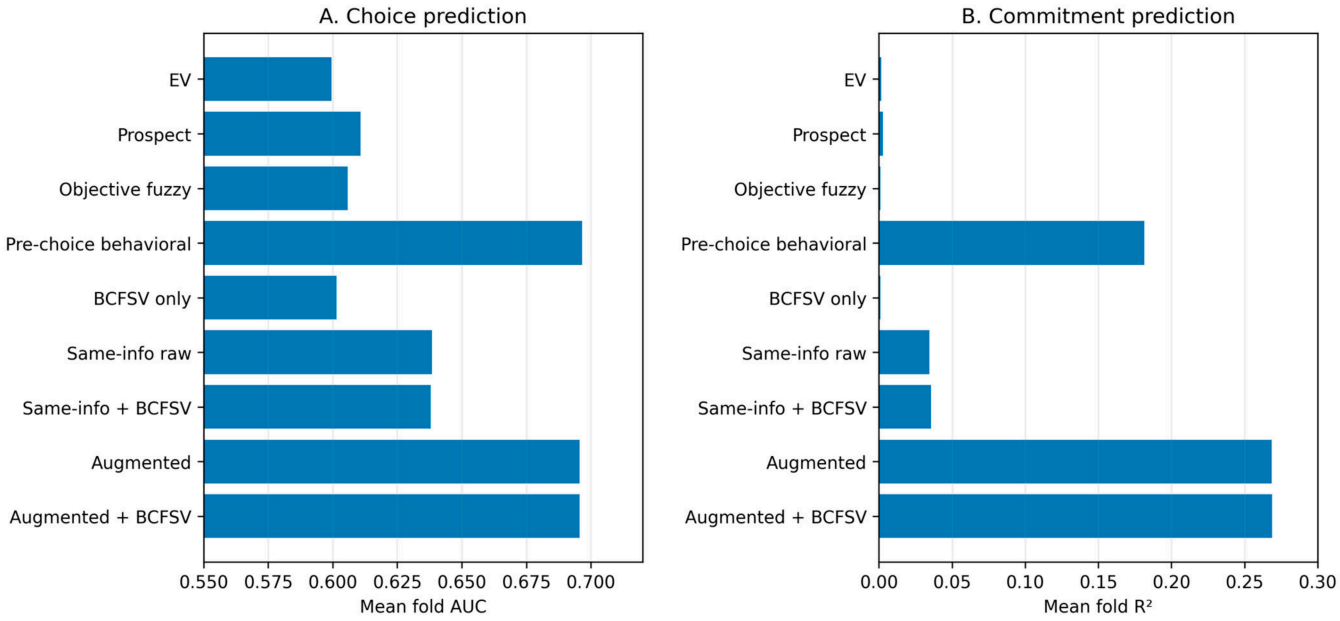


Figure 4. Participant-grouped predictive performance. (**A**) reports mean-fold AUC for choice; (**B**) reports mean-fold R^2 for commitment. The first four models use strict pre-choice information, whereas the remaining models assess held-out-participant generalization with later-measured participant characteristics where applicable.

4.6. Robustness Analyses

The robustness analyses examined whether the focal conclusions depended on calibration scoring, data-quality thresholds, framing or managerial-experience subgroups, control specification, covariance treatment, or the functional form of the BCFSV hesitation attenuation. These checks were interpreted as specification robustness rather than independent replication.

First, H8 remained statistically clear with each alternative calibration score derived from the same 12-task battery. The confidence interaction was 0.908 with numeracy-adjusted Brier error ($SE = 0.125$, $p < 0.001$), 0.742 with overconfidence ($SE = 0.129$, $p < 0.001$), and −0.715 with calibration accuracy ($SE = 0.128$, $p < 0.001$). The negative accuracy interaction is directionally consistent because higher accuracy denotes better calibration, whereas higher Brier error and overconfidence denote poorer calibration.

Second, excluding the fastest approximately 5% of retained respondents reduced the sample to 498 participants and 5976 trials. The focal estimates remained stable: H1 $b = -0.401$, H3 $b = 0.169$, H4 $b = 0.225$, H5 $b = 6.297$, H7 $b = 2.310$, and H8 $b = 0.789$; all $p < 0.001$. Thus, the main behavioral conclusions were not driven by the fastest retained respondents.

Frame-specific analyses showed the same substantive pattern. In gain trials, H1 was −0.501 and H8 was 0.840; in loss trials, H1 was −0.293 and H8 was 0.962; all four estimates were significant at $p < 0.001$. The H1 and H8 findings therefore appeared within both framing conditions rather than being confined to one frame.

The results also remained stable among the 476 formal managers (H1 = −0.412; H8 = 0.916; both $p < 0.001$) and across managerial-experience groups. For participants with six or fewer years of managerial experience, H1 = −0.385 and H8 = 0.862; for those with more than six years, H1 = −0.408 and H8 = 0.887; all $p < 0.001$.

Adding age as an additional participant-level control had negligible influence on the commitment results: H5 remained 6.361, H7 remained 2.335, and H8 remained 0.873 (all $p < 0.001$). The focal commitment findings therefore did not depend on excluding age from the primary parsimonious control set.

Alternative covariance treatment also preserved the interaction conclusions. With two-way participant/scenario clustered inference, the H3 ambiguity × intolerance-of-uncertainty estimate was 0.181 ($SE = 0.020$, $z = 9.28$) and the H8 confidence × Raw Brier estimate was 0.917 ($SE = 0.152$, $z = 6.03$). These checks support the stability of the conclusions to the dependence specification while leaving the primary random-effects structures unchanged.

The H9 null conclusion was also insensitive to the hesitation-attenuation function. Exponential, reciprocal, linear, and quadratic-exponential transformations all yielded same-information choice AUC values of approximately 0.638 and commitment R^2 values of approximately 0.036, with negligible changes relative to the raw-input benchmark. Varying θ from 0.25 to 1.00 likewise did not create an incremental predictive advantage; same-information choice AUC remained approximately 0.637 and commitment R^2 approximately 0.038 across the tested values.

Finally, uncertainty in the pilot BCFSV coefficients was propagated through 2000 draws from the participant-cluster-robust covariance matrix. The main-sample external correlation between predicted π and normalized perceived ambiguity remained within approximately 0.605–0.625 across the central 95% of simulations. This result supports the stability of the criterion-validation conclusion without suggesting that the BCFSV transformation itself has incremental predictive superiority.

Taken together, the behavioral findings were robust to alternative calibration scoring, stricter quality screening, frame and managerial-experience subgroups, age adjustment, and alternative covariance treatment. The model-validation conclusions were likewise stable to coefficient uncertainty and alternative hesitation attenuation. None of these checks altered the central predictive result that H9 was not supported.

Table 10 summarizes the robustness analyses and their implications for the study conclusions.

Table 10. Summary of robustness analyses.

Robustness Test	Alternative Specification	Key Result	Conclusion
Calibration	Adjusted Brier	H8 $b = 0.908$, $SE = 0.125$, $p < 0.001$	Stable
Calibration	Overconfidence	H8 $b = 0.742$, $SE = 0.129$, $p < 0.001$	Stable
Calibration	Accuracy	H8 $b = -0.715$, $SE = 0.128$, $p < 0.001$	Stable; reverse sign expected
Stricter quality screen	Exclude fastest ~5%; $N = 498$	H1 = −0.401; H3 = 0.169; H4 = 0.225; H5 = 6.297; H7 = 2.310; H8 = 0.789; all $p < 0.001$	Stable
Frame subsamples	Gain/loss	Gain: H1 = −0.501, H8 = 0.840; Loss: H1 = −0.293, H8 = 0.962; all $p < 0.001$	Stable
Formal managers	$n = 476$	H1 = −0.412; H8 = 0.916; both $p < 0.001$	Stable
Managerial experience	≤6/>6 years	≤6: H1 = −0.385, H8 = 0.862; >6: H1 = −0.408, H8 = 0.887; all $p < 0.001$	Stable

Table 10. *Cont.*

Robustness Test	Alternative Specification	Key Result	Conclusion
Age control	Add age to commitment models	H5 = 6.361; H7 = 2.335; H8 = 0.873; all $p < 0.001$	Stable
Covariance treatment	Two-way participant/scenario clustered SEs	H3 = 0.181 ($SE = 0.020$); H8 = 0.917 ($SE = 0.152$)	Stable
BCFSV functional sensitivity	Attenuation forms; $\theta = 0.25$–1.00	Choice AUC ≈ 0.637–0.638; commitment $R^2 \approx 0.036$–0.038; no incremental gain	H9 null unchanged

The robustness results therefore strengthen the interpretation of H1–H8 while preserving appropriate caution around the calibration battery's modest person-level precision. They also reinforce the separation between empirical calibration of the BCFSV hesitation representation and incremental predictive value: the former was supported by independent-pilot and external criterion checks, whereas the latter was not detected once the same underlying information was supplied directly to the benchmark models.

5. Discussion

5.1. Main Findings

The findings show that managerial responses to ambiguity depend on both the informational environment and the way that environment is appraised. Ambiguous probability information lowered the pre-choice subjective value of the focal uncertain option (H1), and ambiguity was associated with lower value through higher perceived ambiguity (H2). The participant-cluster bootstrap indirect effect was clear, whereas the residual direct effect was small and nonsignificant after perceived ambiguity was included. This pattern is consistent with an appraisal account of ambiguity, but because perceived ambiguity was measured rather than randomized, it is interpreted as a statistical indirect effect rather than as proof of a causal mediating mechanism. More broadly, the results support treating ambiguity as distinct from simply a more extreme form of measurable risk, consistent with evidence that responses to ambiguity vary with its degree, structure, and subjective interpretation (Bühren et al., 2023; Klingebiel & Zhu, 2023; Aydogan et al., 2026).

The effects of ambiguity also varied systematically across decision makers and frames. Intolerance of uncertainty strengthened the relationship between objective ambiguity and perceived ambiguity (H3): participants higher in intolerance of uncertainty showed larger ambiguity-related increases in perceived ambiguity. This result is consistent with research linking intolerance of uncertainty to indecisiveness, information seeking, and safety-oriented behavior (Appel et al., 2024; Bartoszek et al., 2022), while also supporting a qualified interpretation in which intolerance of uncertainty is especially relevant when the precision or reliability of probability information is itself uncertain (Cobos et al., 2024). Assigned gain/loss framing also moderated the ambiguity–valuation relationship (H4), with a larger ambiguity-related reduction in subjective value in gain-framed than loss-framed trials. Because participant-level frame-recognition responses were not retained, this finding concerns the assigned framing condition rather than conscious recognition of the frame. The pattern is consistent with evidence that framing effects depend on task and contextual features (Balart et al., 2022; Ding et al., 2025).

The commitment findings extend this appraisal sequence beyond valuation. The higher of the two option-attractiveness ratings recorded before choice—BestPreChoiceValue—was strongly associated with later decision commitment (H5). Using this choice-independent measure also avoids conditioning pre-choice value on the option selected later. Perceived ambiguity was negatively associated with BestPreChoiceValue, which in turn was positively

associated with commitment, producing a significant participant-cluster-bootstrap indirect effect (H6). The residual association between perceived ambiguity and commitment remained negative, indicating that pre-choice value captured an important but incomplete part of the ambiguity–commitment relationship. As with H2, H6 is interpreted as a statistical indirect effect because the mediator was not experimentally manipulated.

A second central finding concerns the distinction between confidence and calibrated confidence. Decision confidence was positively associated with commitment even after pre-choice value and the experimental factors were taken into account (H7). More importantly, the confidence–commitment association was stronger among participants with poorer calibration on the objective task battery (H8). Using raw Brier error as the primary calibration measure, the estimated confidence slope increased from 1.48 commitment points per 10 confidence points at relatively better calibration to 3.23 at relatively poorer calibration. The same directional pattern was obtained with numeracy-adjusted Brier error, overconfidence, and calibration accuracy. Thus, high expressed confidence should not be equated automatically with better judgment; in this dataset, confidence was more strongly linked to commitment among participants whose confidence was less well aligned with objective performance. This distinction is consistent with views of confidence as a noisy estimate of decision reliability (Boundy-Singer et al., 2023; Fleming, 2024) and with research on overconfidence in consequential decisions (Boyle et al., 2025; Boyle & Nye, 2026).

The BCFSV analyses produced a deliberately bounded methodological conclusion. The hesitation component was estimated in the independent pilot against observed normalized perceived ambiguity and then transported to the main sample without refitting to choice or commitment. The pilot mapping showed moderate external correspondence with the same criterion in the main sample, supporting its use as an empirically anchored ambiguity-response representation. However, this calibration success did not translate into incremental predictive superiority. Strict pre-choice behavioral information predicted choice and later commitment better than the simpler expected-value, prospect-value, and objective-fuzzy representations, whereas adding the nonlinear BCFSV transformation to an exact same-information benchmark did not produce detectable improvement in AUC, error metrics, or out-of-sample R^2. H9 was therefore not supported. The results distinguish two questions that should not be conflated: whether a fuzzy hesitation representation can be empirically calibrated, and whether transforming the underlying information through that representation adds prediction beyond entering the same information directly.

5.2. Theoretical Implications

The study contributes to behavioral research on ambiguity by linking objective ambiguity, subjective appraisal, and valuation within the same repeated-decision design. The clear ambiguity → perceived ambiguity → subjective-value indirect effect is consistent with the proposition that the psychological experience of imprecise probability information is closely involved in how ambiguous alternatives are valued. At the same time, the study does not treat perceived ambiguity as a causally identified mediator. This distinction matters theoretically because heterogeneity in ambiguity responses need not imply a single stable preference for or against ambiguity; it may partly reflect how the same informational structure is experienced by different decision makers. The moderation by intolerance of uncertainty reinforces this interpretation and complements evidence of substantial heterogeneity in ambiguity aversion (Klingebiel & Zhu, 2023; Aydogan et al., 2026). The framing result further suggests that this appraisal process is reference-sensitive: ambiguity affected valuation under both frames, but its magnitude differed depending on whether the decision was presented around gains or losses.

The commitment results add a second theoretical layer by separating pre-choice value, confidence, and confidence calibration. BestPreChoiceValue was strongly associated with commitment and statistically carried part of the association between perceived ambiguity and later commitment, while confidence contributed additional information beyond pre-choice value. Calibration then qualified the meaning of confidence: the confidence–commitment relationship was stronger when objective calibration was poorer. This pattern helps reconcile two views of managerial confidence—as useful decisiveness and as potentially misleading overconfidence—by indicating that the behavioral meaning of confidence depends partly on how well it corresponds to objective performance. Two managers expressing the same degree of confidence may differ in how accurately that confidence reflects their performance on objectively scored decisions. When confidence is closely aligned with performance, it may provide a useful signal of judgment reliability; when calibration is poorer, similar expressed certainty may accompany stronger commitment without comparable evidence of accurate judgment. In the present study, the confidence–commitment association was stronger at higher Raw Brier error, consistent with this distinction. Because calibration was measured using a separate task battery and the findings are associational, this pattern should not be interpreted as proof that poorly calibrated confidence causes excessive managerial commitment. It is consistent with contextual accounts of subjective value (Zheng et al., 2026; Soukupova et al., 2026) and with management research linking overconfidence to strategic risk taking and organizational outcomes (Brunzel, 2021; Burkhard et al., 2023; Kunz & Sonnenholzner, 2023). Taken together, the findings support a person-by-context account of managerial judgment in which objective uncertainty, subjective appraisal, reference framing, pre-choice valuation, and metacognitive calibration jointly shape how uncertainty is translated into commitment.

5.3. Methodological Implications

The methodological contribution of the BCFSV framework is not the mere combination of fuzzy logic and prospect-based valuation; related hybrid approaches are already established in decision analysis (Tian et al., 2022; Pan et al., 2024; Wu et al., 2024), and personalization by itself is not sufficient as a novelty claim given recent interpretable and co-constructed fuzzy-set methods (García-Zamora et al., 2024; Dutta et al., 2026). The more specific contribution is the use of an independent pilot to empirically calibrate the hesitation term against an observed pre-choice ambiguity appraisal. In the specification, π is a pilot-calibrated ambiguity-response representation based on objective ambiguity, ambiguity width, pilot-standardized intolerance of uncertainty, and the ambiguity × intolerance-of-uncertainty interaction. Calibration error is not forced into this equation; it remains a separate construct for the confidence-calibration hypothesis.

The analysis also clarifies the roles of the fuzzy components. The primary BCFSV score uses the pilot-calibrated hesitation term π to attenuate the signed prospect-value differential. The support and opposition components, μ and ν, are algebraically derived from π and normalized pre-choice subjective value and are therefore interpretive decompositions rather than independently estimated or separately validated predictors. This distinction prevents the fuzzy architecture from appearing to contain more empirical degrees of freedom than were actually estimated. The prospect-value component is likewise fully reproducible: its curvature, loss-aversion, probability-weighting, monetary-scaling, and gain/loss reference-point rules are explicitly documented and reproduce all archived prospect-value scores to the recorded rounding precision.

Finally, the predictive design provides a methodological boundary test rather than a demonstration exercise. Strict pre-choice prediction was separated from held-out-participant generalization because intolerance of uncertainty and calibration-related par-

ticipant information were collected after the 12 managerial trials. Within the held-out-participant analyses, BCFSV was compared with an exact same-information benchmark containing its raw inputs, and paired participant-bootstrap intervals were used to test the incremental metric differences. The absence of detectable improvement means that the value of BCFSV in the present data is representational and interpretive, not evidence that nonlinear fuzzy transformation is inherently more predictive. Future methodological work should evaluate whether prospectively available individual-difference histories, richer repeated calibration data, alternative empirically justified hesitation functions, or genuinely new information sources produce incremental value beyond matched raw-input models.

5.4. Managerial Implications

The clearest managerial implication concerns confidence. Expressed confidence should not be treated as a direct proxy for decision quality, because confidence was more strongly associated with commitment among participants with poorer objective calibration. For important forecasts, investment recommendations, project approvals, or risk assessments, organizations could record both the manager's confidence and the eventual verifiable outcome. Repeated confidence–performance feedback can help distinguish consistently informative confidence from persistent overconfidence or underconfidence without treating every unsuccessful decision as poor judgment under uncertainty.

The ambiguity findings also provide a behavioral rationale for communicating uncertainty more explicitly in managerial decision materials. Ambiguous probability information reduced the subjective value assigned to the uncertain option (H1), and perceived ambiguity was involved in this relationship through a significant statistical indirect effect (H2). Moreover, intolerance of uncertainty strengthened the association between objective and perceived ambiguity (H3), indicating that the same informational conditions can be appraised differently across decision makers. One potential implication is that decision materials could distinguish precise risk estimates from probability ranges when the evidence does not justify a single point estimate and explain why a range is wide. Structured reviews could also ask managers to distinguish the available probability evidence, their subjective valuation, and their confidence in a recommendation. This is a potential practical implication of the observed appraisal patterns, not evidence that the present experiment directly tested or established the effectiveness of uncertainty-communication interventions. Making ambiguity visible should not be taken as a reason to reject uncertain opportunities automatically.

The framing result suggests an additional practical safeguard. Economically similar decisions can be described as opportunities to achieve gains or as efforts to avoid shortfalls, and the magnitude of the ambiguity effect differed across these assigned frames. For major proposals, boards, senior leaders, and analysts could therefore examine whether the recommendation remains stable when the same underlying information is presented from both gain- and loss-oriented reference points. Post-decision reviews can then focus not only on whether an outcome was favorable, but also on whether the uncertainty, valuation, and confidence expressed at the time were proportionate to the information then available.

BCFSV is better suited to this type of diagnostic reflection than to automated managerial ranking. The present study supports an empirically calibrated representation of ambiguity response, but it does not show incremental predictive superiority over models containing the same information directly, and the participant-level IU information used by BCFSV was collected after the focal managerial trials. A practical system inspired by the framework should therefore display uncertainty, appraisal, and historical calibration as separate decision-audit inputs rather than convert them into an unquestioned decision score. Prospective organizational use would require those person-level histories to be available before the focal decision and to be validated against consequential outcomes.

5.5. Limitations and Future Research

Several limitations bound the interpretation of the findings. First, the study relied on scenario-based managerial decisions without actual financial consequences. This design made it possible to control ambiguity, framing, probability information, and monetary quantities, but it cannot reproduce the incentives, accountability pressures, organizational politics, resource constraints, or reputational consequences of real managerial choices. The 12 scenarios also represent a fixed and necessarily limited bank of strategic and operational contexts. The primary mixed models therefore treat scenario as a fixed experimental factor, and the results should not be interpreted as unrestricted generalization to all managerial decisions. Future studies should use financially incentivized tasks, field experiments, industry-specific settings, and observed high-stakes decisions to test whether the same relationships hold under consequential conditions.

Second, confidence calibration was measured with a separate 12-task objective battery because the managerial ambiguity scenarios did not have objectively correct outcomes. Raw Brier error was used as the primary H8 calibration indicator, while numeracy-adjusted Brier error, overconfidence, and accuracy served as sensitivity measures. The convergence across these scoring approaches supports the direction of the H8 finding, but all derive from the same task battery and therefore do not constitute independent replication. Person-level precision was modest: odd-even Spearman–Brown coefficients were 0.594 for accuracy, 0.422 for raw Brier error, 0.385 for numeracy-adjusted Brier error, and 0.371 for overconfidence. Calibration-related effect sizes should therefore be interpreted conservatively. Future work could improve precision through larger task banks, repeated sessions, domain-matched calibration tasks, and longitudinal feedback linked to realized outcomes. The three verified numeracy items are reported in Table S3, but numeracy remains an ancillary adjustment variable rather than a focal validated construct.

Third, the indirect-effect analyses do not identify causal mediation. H2 preserves a meaningful temporal sequence from experimentally assigned ambiguity to perceived ambiguity and then to pre-choice valuation, but perceived ambiguity itself was not randomized. H6 uses BestPreChoiceValue, which is defined entirely before choice and is not conditioned on the later selected option; nevertheless, neither perceived ambiguity nor BestPreChoiceValue was experimentally manipulated. The H2 and H6 results should therefore be read as statistical indirect effects. Designs that manipulate appraisals or valuation-relevant information, add greater temporal separation, or use assumptions and estimators developed specifically for causal mediation would provide stronger evidence about mechanism.

Fourth, the framing analysis is limited by the absence of retained participant-level frame-recognition responses. Gain and loss variants were numerically matched within each scenario, and expert ratings supported the clarity of the framing distinction, but the data cannot establish that every participant consciously recognized the assigned frame on every trial. H4 is therefore an interaction with the experimentally assigned framing condition, not a participant-recognition effect. Future studies should retain trial-level recognition responses and test whether framing effects differ according to explicit frame awareness.

Fifth, the sample consisted of organizational professionals in Türkiye. Responses to ambiguity, confidence, framing, and commitment may vary with national institutions, organizational cultures, accountability norms, and accepted ways of expressing certainty. Cross-national and cross-industry replication is needed before treating the observed relationships as culturally general. Such work would be especially useful for determining whether the moderation by intolerance of uncertainty and the confidence-calibration pattern operate similarly across managerial cultures with different norms around risk and decisiveness.

Sixth, the BCFSV calibration should be interpreted within its empirical scope. The hesitation term was calibrated in an independent pilot of 42 participants against normal-

ized perceived ambiguity; it was not calibrated against a separately administered hesitation scale. The resulting π therefore represents one empirically anchored ambiguity-response operationalization rather than a uniquely established psychological hesitation construct. Although participant-grouped, scenario-held-out, external-main-sample, coefficient-uncertainty, attenuation-function, and θ sensitivity checks were stable, other theoretically defensible fuzzy formulations could represent uncertainty differently. In addition, intolerance of uncertainty was measured after the 12 managerial scenarios. Consequently, the BCFSV analyses establish held-out-participant generalization conditional on later-measured participant information, not real-time prospective prediction at the focal choice. Prospective validation would require IU and any calibration history to be available before decisions occur.

Finally, H9 was not supported: the nonlinear BCFSV transformation did not add detectable predictive information beyond an exact same-information benchmark or a richer behavioral benchmark. This null result defines an important boundary rather than a reason to tune the model after the fact. Future research should test BCFSV in longitudinal organizational settings where individual calibration histories are accumulated before decisions, where senior managers make repeated consequential choices, and where realized outcomes permit direct evaluation of both calibration and decision quality. Combining cross-national samples, incentivized or field decisions, repeated measures, and prospective person-level histories would provide a stronger basis for assessing external validity and determining when, if ever, behaviorally calibrated fuzzy transformations add explanatory or predictive value beyond the same information entered directly.

6. Conclusions

This study shows that managerial responses to ambiguous information are closely related to how that information is appraised and valued. Ambiguous probability information reduced the pre-choice subjective value of the uncertain option, and a participant-cluster-bootstrap indirect effect linked objective ambiguity to lower value through higher perceived ambiguity. Because perceived ambiguity was measured rather than randomized, this result is best understood as a statistical indirect pathway, not as definitive causal mediation.

The findings also identify meaningful person- and context-level differences. Intolerance of uncertainty strengthened the ambiguity–perceived ambiguity relationship, and assigned gain/loss framing changed the magnitude of the ambiguity–valuation effect. Pre-choice value was strongly associated with later commitment, and the choice-independent BestPreChoiceValue construct statistically carried part of the association between perceived ambiguity and commitment. These results support a view of managerial valuation as jointly shaped by objective uncertainty, subjective appraisal, individual differences, and reference context.

Confidence added a further metacognitive dimension. Higher decision confidence was associated with stronger commitment, but this association was substantially stronger among participants with poorer objective calibration. Raw Brier error provided the primary evidence for this moderation, and the same substantive pattern appeared with alternative calibration scores. The implication is not that confidence is undesirable, but that its behavioral meaning depends partly on whether expressed certainty has been well calibrated to objective performance.

Methodologically, the BCFSV framework provides an empirically anchored way to represent person-specific ambiguity response. Its hesitation component was calibrated in an independent pilot against observed perceived ambiguity and transported to the main sample without refitting to choice or commitment; the prospect-value component was also made fully reproducible through explicit parameter and reference-point documentation. The predictive analyses nevertheless set a clear boundary: behavioral appraisal and valuation

information improved prediction relative to simpler value-only benchmarks, but the BCFSV transformation did not provide detectable incremental predictive gain beyond models containing the same underlying information. H9 was therefore not supported, and BCFSV should be interpreted primarily as a structured, transparent representation rather than as a superior prediction engine.

The broader contribution is to shift attention from uncertainty as a property of the decision alone toward the interaction between the decision environment and the manager interpreting it. For research, the next step is prospective, cross-national, and consequential validation using richer calibration histories and realized outcomes. For practice, the findings support decision processes that make ambiguity, subjective appraisal, confidence, and historical calibration visible without allowing any single score to substitute for managerial judgment.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/bs16112021/s1, The Supporting Information includes Table S1. Supporting Information Index and Documentation Notes; Table S2. Variable Dictionary, Coding, and Scoring; Table S3. Survey Instrument and Administration Sequence; Table S4. Scenario Bank and Counterbalancing; Table S5. Objective Confidence-Calibration Tasks; Table S6. Analysis and Reproducibility Details; Table S7. Measurement Sources and Documentation; Table S8. Raw Recruitment and Screening Data ($N = 560$); Table S9. Main-Sample Participant-Level Data ($N = 525$); Table S10. Main-Study Trial-Level Data ($N = 6300$); Table S11. Confidence-Calibration Task Bank; Table S12. Confidence-Calibration Task Quality; Table S13. Main-Sample Calibration Responses ($N = 6300$); Table S14. Expert Ratings ($N = 7$); Table S15. Expert Content Validity Index Summary; Table S16. Pilot-Sample Participant Characteristics ($N = 42$); Table S17. Pilot Trial-Level Data ($N = 504$); Table S18. Pilot Calibration Responses ($N = 504$); Table S19. Scenario Design and Ambiguity Parameters; Table S20. Methodological Documentation; Table S21. Supplementary Results Summary; Table S22. Analysis Specification; Table S23. Scenario-Level Choice Rates; and Table S24. Pilot Model Provenance and Specification Timing. These materials provide the study instrument, raw and derived data, calibration-task information, validation evidence, methodological documentation, analysis specifications, robustness information, and supplementary results needed to support transparency and reproducibility. Table S3 reproduces the verified original NUM1–NUM3 wording, response formats, correct answers, and scoring rules, while Table S7 documents their source/provenance and ancillary analytical role. The following references are cited in the supplementary materials: Arts et al. (2024), Boundy-Singer et al. (2023), Carleton et al. (2007), Dohmen et al. (2011), Hur et al. (2026), Olschewski et al. (2025), Toribio-Flórez et al. (2023), and Valencia et al. (2025).

Author Contributions: Conceptualization, F.M., H.G.D. and T.K.; Methodology, F.M., H.G.D. and T.K.; Software, F.M., H.G.D. and T.K.; Validation, F.M., H.G.D. and T.K.; Formal analysis, F.M., H.G.D. and T.K.; Investigation, F.M., H.G.D. and T.K.; Resources, F.M., H.G.D. and T.K.; Data curation, F.M., H.G.D. and T.K.; Writing—original draft, F.M., H.G.D. and T.K.; Writing—review and editing, F.M., H.G.D. and T.K.; Visualization, F.M., H.G.D. and T.K.; Supervision, F.M.; Project administration, F.M.; Funding acquisition, F.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was conducted in accordance with the ethical principles of the Declaration of Helsinki. Ethical approval was obtained from the Ethics Committee of Istanbul Atlas University. (protocol code E-22686390-050.99-95112. and 13 April 2026 of approval).

Informed Consent Statement: All participants were informed about the purpose of the study, the voluntary nature of participation, confidentiality procedures, and their right to withdraw from the study at any time. Informed consent was obtained from all participants before their involvement in the research. Informed consent was obtained from all participants involved in the study.

Data Availability Statement: The original contributions presented in this study are included in the article and Supporting Information. The de-identified study data used to reproduce the reported analyses are provided as supplementary data accompanying the submission. The Supporting Information

also reproduces the verified original wording and scoring details for the three ancillary numeracy checks. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

Aguilar-Lleyda, D., & de Gardelle, V. (2021). Confidence guides priority between forthcoming tasks. *Scientific Reports*, *11*, 18320. [CrossRef] [PubMed]

Almazrouei, M. A., Dror, I. E., Morgan, R. M., Dan, O., Paterson, M., & Levy, I. (2025). Human factors in triaging forensic items: Casework pressures and ambiguity aversion. *Science & Justice*, *65*(2), 149–162. [CrossRef] [PubMed]

Appel, H., Krasko, J., Luhmann, M., & Gerlach, A. L. (2024). Intolerance of uncertainty predicts indecisiveness and safety behavior in real-life decision making: Results from an experience sampling study. *Journal of Research in Personality*, *110*, 104490. [CrossRef]

Arts, S., Ong, Q., & Qiu, J. (2024). Measuring decision confidence. *Experimental Economics*, *27*, 582–603. [CrossRef]

Ashraf, S., Jana, C., Iqbal, W., & Deveci, M. (2025). Regret-based domination three-way decision-making model with circular spherical fuzzy Mahalanobis distance. *Information Sciences*, *691*, 121616. [CrossRef]

Aydogan, I., Berger, L., & Bosetti, V. (2026). Unraveling ambiguity aversion. *The Review of Economics and Statistics*, *108*(2), 533–541. [CrossRef]

Balart, P., Ezquerra, L., & Hernandez-Arenaz, I. (2022). Framing effects on risk-taking behavior: Evidence from a field experiment in multiple-choice tests. *Experimental Economics*, *25*, 1268–1297. (Correction to 2022, *Experimental Economics*, *25*, 1298). [CrossRef]

Balsdon, T., & Philiastides, M. G. (2024). Confidence control for efficient behaviour in dynamic environments. *Nature Communications*, *15*, 9089. [CrossRef] [PubMed]

Bartoszek, G., Ranney, R. M., Curanovic, I., Costello, S. J., & Behar, E. (2022). Intolerance of uncertainty and information-seeking behavior: Experimental manipulation of threat relevance. *Behaviour Research and Therapy*, *154*, 104125. [CrossRef] [PubMed]

Boundy-Singer, Z. M., Ziemba, C. M., & Goris, R. L. T. (2023). Confidence reflects a noisy decision reliability estimate. *Nature Human Behaviour*, *7*(1), 142–154. [CrossRef] [PubMed]

Boyle, P. J., & Nye, P. (2026). Information distortion as a source of overconfidence in managerial decisions. *Judgment and Decision Making*, *21*, e3. [CrossRef]

Boyle, P. J., Russo, J. E., & Kim, J. (2025). When deciding creates overconfidence. *Judgment and Decision Making*, *20*, e15. [CrossRef]

Brunzel, J. (2021). Overconfidence and narcissism among the upper echelons: A systematic literature review. *Management Review Quarterly*, *71*, 585–623. [CrossRef]

Burkhard, B., Sirén, C., van Essen, M., Grichnik, D., & Shepherd, D. A. (2023). Nothing ventured, nothing gained: A meta-analysis of CEO overconfidence, strategic risk taking, and performance. *Journal of Management*, *49*(8), 2629–2666. [CrossRef]

Bühren, C., Meier, F., & Pleßner, M. (2023). Ambiguity aversion: Bibliometric analysis and literature review of the last 60 years. *Management Review Quarterly*, *73*, 495–525. (Correction to 2022, *Management Review Quarterly*, *72*, 1287). [CrossRef]

Carleton, R. N., Norton, M. A. P. J., & Asmundson, G. J. G. (2007). Fearing the unknown: A short version of the intolerance of uncertainty scale. *Journal of Anxiety Disorders*, *21*(1), 105–117. [CrossRef] [PubMed]

Chen, D. L. (2026). Testing axiomatizations of ambiguity aversion. *Journal of Behavioral and Experimental Finance*, *49*, 101134. [CrossRef]

Chochoiek, N., Huber, L. R., & Sloof, R. (2025). Optimism and overconfidence of strategic decision makers—Comparing entrepreneurs and managers with employees. *Journal of Economics & Management Strategy*, *34*(3), 674–695. [CrossRef]

Cobos, P. L., Quintero, M. J., López, F. J., Luque, D., Ciria, L. F., & Morís, J. (2024). Intolerance of uncertainty does not significantly predict decisions about delayed, probabilistic rewards: A failure to replicate Luhmann, C. C., Ishida, K., & Hajcak, G. (2011). *PLoS ONE*, *19*(7), e0298503. [CrossRef] [PubMed]

Cohee, G. L., & Barnhart, C. M. (2024). Often wrong, never in doubt: Mitigating leadership overconfidence in decision-making. *Organizational Dynamics*, *53*(3), 101011. [CrossRef]

Ding, X., Li, M., & Qiao, J. (2025). Meta-analysis examining the relationship between framing effect and risky decisions. *Journal of Behavioral and Experimental Economics*, *116*, 102351. [CrossRef]

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, *9*(3), 522–550. [CrossRef]

Double, K. S., & Birney, D. P. (2024). Confidence judgments interfere with perceptual decision making. *Scientific Reports*, *14*, 14133. [CrossRef] [PubMed]

Dutta, B., García-Zamora, D., Figueira, J. R., & Martínez, L. (2026). A deck of cards-based co-constructive approach for modeling higher-order uncertainty in fuzzy decision-making. *Applied Soft Computing*, *196*, 115094. [CrossRef]

Eti, S., Ergün, E., Acar, M., Yüksel, S., & Dinçer, H. (2026). Prioritizing psychological factors in young consumers' coffee brand preferences through a novel behavioral leadership fuzzy decision-making modelling. *Scientific Reports*, *16*, 24263. [CrossRef] [PubMed]

Fitz, E. B., Stecuła, D. A., Hitt, M. P., & Saunders, K. L. (2024). Objective numeracy exacerbates framing effects from decision-making under political risk. *Scientific Reports*, *14*, 10473. [CrossRef] [PubMed]

Fleming, S. M. (2024). Metacognition and confidence: A review and synthesis. *Annual Review of Psychology*, *75*, 241–268. [CrossRef] [PubMed]

Gaba, V., Lee, S., Meyer-Doyle, P., & Zhao-Ding, A. (2023). Prior experience of managers and maladaptive responses to performance feedback: Evidence from mutual funds. *Organization Science*, *34*(2), 894–915. [CrossRef]

García-Zamora, D., Dutta, B., & Figueira, J. R. (2026). Consensus-based deck of cards method for constructing fuzzy linguistic scales. *Group Decision and Negotiation*, *35*, 57. [CrossRef]

García-Zamora, D., Dutta, B., Figueira, J. R., & Martínez, L. (2024). The deck of cards method to build interpretable fuzzy sets in decision-making. *European Journal of Operational Research*, *319*(1), 246–262. [CrossRef]

Hsee, C. K., Imas, A., & Li, X. (2024). How framing influences strategic interactions. *Management Science*, *71*(4), 3013–3025. [CrossRef]

Hu, Y., Lu, X., Zheng, W., Wang, L., & Yu, P. (2023). The neurobase of ambiguity loss aversion about decision making. *Frontiers in Psychology*, *14*, 1055640. [CrossRef] [PubMed]

Hur, E. Y., Woolley, K., & Tu, Y. (2026). Converting preference to actual commitment: The role of option presentation order. *Journal of the Academy of Marketing Science*, *54*, 948–967. [CrossRef]

Jiroušek, R., & Kratochvíl, V. (2020). On subjective expected value under ambiguity. *International Journal of Approximate Reasoning*, *127*, 70–82. [CrossRef]

Klingebiel, R., & Zhu, F. (2023). Ambiguity aversion and the degree of ambiguity. *Journal of Risk and Uncertainty*, *67*(3), 299–324. [CrossRef]

Kohl, A. T., Sauer, J. D., Palmer, M. A., Brooks, J., & Heathcote, A. (2024). The effects of non-diagnostic information on confidence and decision making. *Memory & Cognition*, *52*(5), 1182–1194. [CrossRef] [PubMed]

Krypotos, A. M., Alves, M., Crombez, G., & Vlaeyen, J. W. S. (2022). The role of intolerance of uncertainty when solving the exploration-exploitation dilemma. *International Journal of Psychophysiology*, *181*, 33–39. [CrossRef] [PubMed]

Kunz, J., & Sonnenholzner, L. (2023). Managerial overconfidence: Promoter of or obstacle to organizational resilience? *Review of Managerial Science*, *17*, 67–128. [CrossRef]

Kuzmics, C., Rogers, B. W., & Zhang, X. (2024). Randomization advice and ambiguity aversion. *Journal of Risk and Uncertainty*, *69*(1), 85–104. [CrossRef]

Leder, J., Chapkovski, P., Schütz, A., Lauer, T., & Gürerk, Ö. (2024). Background uncertainty does not increase risk aversion in decision making. *Scientific Reports*, *14*, 25899. [CrossRef] [PubMed]

Olschewski, S., Mullett, T. L., & Stewart, N. (2025). Optimal allocation of time in risky choices under opportunity costs. *Cognitive Psychology*, *157*, 101716. [CrossRef] [PubMed]

Pan, X.-H., He, S.-F., & Wang, Y.-M. (2024). A new decision analysis framework for multi-attribute decision-making under interval uncertainty. *Fuzzy Sets and Systems*, *480*, 108867. [CrossRef]

Sakamoto, Y., & Miyoshi, K. (2024). A confidence framing effect: Flexible use of evidence in metacognitive monitoring. *Consciousness and Cognition*, *118*, 103636. [CrossRef] [PubMed]

Schneck, S., & Hautz, J. (2024). Overconfident executives: Do they wreak havoc or inspire the troops? *Journal of Business Strategy*, *45*(2), 115–123. [CrossRef]

Soukupova, M., Garcia, B., & Palminteri, S. (2026). Context induces distortions in value representations across multiple elicitation methods and learning modalities. *Nature Communications*, *17*, 6169. [CrossRef] [PubMed]

Tian, X., Ma, J., Li, L., Xu, Z., & Tang, M. (2022). Development of prospect theory in decision making with different types of fuzzy sets: A state-of-the-art literature review. *Information Sciences*, *615*, 504–528. [CrossRef]

Toribio-Flórez, D., Saße, J., & Baumert, A. (2023). "Proof under reasonable doubt": Ambiguity of the norm violation as boundary condition of third-party punishment. *Personality and Social Psychology Bulletin*, *49*(3), 429–446. [CrossRef] [PubMed]

Treiman, L. S., & Kool, W. (2024). Choosing the right frame: How context preferences facilitate subsequent decisions. *Scientific Reports*, *14*, 31607. [CrossRef] [PubMed]

Valencia, P. D., De la Rosa-Gómez, A., Suzán-Montoya, G., Alvarez-López, A., Pérez-Martínez, Z., & Hernández-Posadas, A. (2025). The IUS-4: A psychometric evaluation of an ultra-brief intolerance of uncertainty measure. *Journal of Rational-Emotive & Cognitive-Behavior Therapy*, *43*, 61. [CrossRef]

Wu, M., Ma, S., & Fan, J. (2024). A spherical Z-number multi-attribute group decision making model based on the prospect theory and GLDS method. *Complex & Intelligent Systems*, *10*, 8501–8524. [CrossRef]

Yalvaç, H. A., Eti, S., Dinçer, H., & Yüksel, S. (2026). Ranking work–family conflict management strategies through a novel multi-criteria decision-making model. *Acta Psychologica*, *267*, 107112. [CrossRef] [PubMed]

Zhang, J. (2022). More ambiguity aversion or more risk aversion? *Economic Theory Bulletin*, *10*(2), 217–232. [CrossRef]

Zhang, Y., Xu, Z., Hao, Z., & Liao, H. (2021). Dynamic assessment of Internet public opinions based on the probabilistic linguistic Bayesian network and prospect theory. *Applied Soft Computing*, *106*, 107359. [CrossRef]

Zhao, S., Ye, B., Wang, W., & Zeng, Y. (2022). The intolerance of uncertainty and "untact" buying behavior: The mediating role of the perceived risk of COVID-19 variants and protection motivation. *Frontiers in Psychology*, *13*, 807331. [CrossRef] [PubMed]
Zheng, M., Ma, Z., Li, Y., & Liang, D. (2026). Subjective values in decision-making under uncertainty: A systematic review. *Behavioral Sciences*, *16*(9), 1494. [CrossRef] [PubMed]

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.