

Article

Formal Inconsistencies of Expertise Aggregation Techniques Commonly Employed in Engineering Teams

Cynthia Stephen ¹, Hanumanthrao Kannan ² and Alejandro Salado ^{3,*} ¹ Grado Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA 24061, USA² Department of Industrial & Systems Engineering and Engineering Management, University of Alabama in Huntsville, Huntsville, AL 35899, USA; hk0049@uah.edu³ Department of Systems and Industrial Engineering, The University of Arizona, Tucson, AZ 85721, USA

* Correspondence: alejandrosalado@arizona.edu

Abstract: Engineering managers leverage the expertise of engineers in their teams to inform decisions. Engineers may convey their expertise in the form of opinions and/or judgements. Given a decision, it is common to elicit and aggregate the expertise from various engineers to capture a broader set of experiences and knowledge. Establishing an internally and externally consistent aggregation framework is therefore paramount to yield a meaningful aggregation, that is, to make sure that the expertise of each engineer is accounted for reasonably. However, we contend that most de facto aggregation techniques lack such consistency and lead to the inadequate use and aggregation of engineering expertise. In this paper, we investigate the consistency or lack thereof of various expertise aggregation techniques. We derive implications of such inconsistencies and provide recommendations about how they may be overcome. We illustrate our discussion using safety decisions in engineering as a notional case.

Keywords: decision making; engineering management; expertise aggregation



Citation: Stephen, C.; Kannan, H.; Salado, A. Formal Inconsistencies of Expertise Aggregation Techniques Commonly Employed in Engineering Teams. *Systems* **2024**, *12*, 180. <https://doi.org/10.3390/systems12050180>

Academic Editor: Ed Pohl

Received: 1 May 2024

Revised: 13 May 2024

Accepted: 16 May 2024

Published: 18 May 2024



Copyright: © 2024 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Engineers contribute to project decisions in two main ways. First, they generate and characterize alternatives. Second, they identify and characterize uncertain events associated with the potential execution of those activities [1]. When multiple engineers work together on a given decision, the alternatives generated by the various engineers in the team can be simply accumulated. That is, the alternatives can be joined as a larger set of alternatives to be evaluated. The same approach can be used for the events that the different engineers identify (even though, in this case, the process is a bit more sophisticated, as relevance relationships need to be accounted for [2]). However, when various experts provide different characterizations of a specific uncertain event, these characterizations must be aggregated. That is, the set of characterizations needs to be consolidated into a single characterization for that event so that it can be used in the decision model [3]. These characterizations may be provided in the form of an opinion (a belief distribution on an event) or a judgement (a bet on a specific outcome). Such an aggregation process is the focus of this paper.

In the past, several methods have been proposed to aggregate the expertise of subject matter experts (SMEs). These can be broadly classified as behavioral and mathematical [4]. Behavioral methods involve an information exchange and the negotiation of opinions among the SMEs to arrive at a consensus. Mathematical aggregation involves obtaining quantitative values (for the information sought) from the SMEs and applying some statistical averaging technique to combine the values [5]. These expertise aggregation methods have been applied in several domains and contexts [6], and their characteristics, such as robustness, traceability, prediction capabilities, and accuracy, have been studied [4]. However, the literature scarcely provides concrete guidelines in expertise aggregation specific

to engineering decision-making. In fact, as we will show with an example, the protocols and procedures established to guide the decision-making process in engineering contexts do not prescribe, nor guide the adequate selection of expertise aggregation methods. It should be noted that fuzzy rule-based methods are not included in this paper because of their lesser adoption in engineering practice.

Consider, for example, the Safety Risk Management Policies (SRMPs) employed by the Federal Aviation Administration (FAA) and the U.S. Navy [7,8]. These policies indicate that multiple SMEs are to be involved in the decision process for safety risk assessment and acceptance. SMEs are tasked with providing inputs into a safety risk matrix based on their *technical* judgement [3], specifically in terms of the severity of a safety risk and its likelihood of occurrence. While the SRMPs describe the actors that must be consulted in the safety decision process, as well as their responsibilities, they do not prescribe the process that dictates how their inputs must be aggregated. Therefore, freedom is granted to choose the aggregation technique for each particular safety risk assessment scenario. No prescription is provided for validating the adequacy of such a selection either. In fact, a report published by the United States General Accounting Office states the following [9]:

“Both, the FAA and the military services have internal informal networks in place among aviation safety personnel to share information. These exchanges are typically self-initiated, occur on an ad hoc basis, and are based largely on personal relationships. It is a primary means used among the military services’ safety centers to keep apprised of current aviation safety issues”.

When a concrete procedure for expertise aggregation is not defined, we suggest that inconsistency-prone behavioral aggregation methods that use expert judgement are likely to be adopted due to their simplicity and low computational complexity, when compared to those of mathematical methods. For example, FAA Orders 8040.4 B and 8040.6 and the U.S. Navy’s OPNAVINST 3500.39 C indicate this trend by using terms such as “mutually agreed upon”, “all stakeholders have been consulted”, “peer reviewed”, or “brainstorming” to describe the aggregation process. As a result, an inadequate aggregation of expertise that may yield an inappropriate selection of mitigation/control strategies would still be used because they would still be seemingly compliant with the regulatory guidelines imposed by the organization.

However, the selection of an aggregation approach is not trivial. The aggregation process is sensitive to the type of data provided by the experts (e.g., probability distributions, qualitative statements, arbitrary scores, etc.) that is being aggregated. Judgement aggregation, for example, is a problem for any group that tries to conclude a judgement based on sets of rationally interconnected propositions [10]. In these cases, consistency problems, such as the impossibility theorem or the discursive dilemma (also known as the doctrinal paradox) [11], might easily emerge and remain undetected. Furthermore, the quality of aggregating expertise depends on factors such as the competence of the expert, the expert team composition, the method of expert elicitation, and the behavioral/mathematical process of aggregation. While these problems have been traditionally associated with issues in economics, social choice, and political theories, they have not been addressed in engineering contexts. However, these factors must be taken into account to choose an adequate aggregation method that achieves a consistent aggregation.

To understand the implications of these aspects, we explore in this paper the ability of different expertise aggregation methods to preserve the premise and judgement consistency, contextualized in engineering settings. While we use an engineering safety decision as a notional case, the arguments presented in this paper are valid for any engineering decision that involves the aggregation of expertise.

This paper is organized as follows. First, we summarize the existing work on problems and potential solutions associated with the aggregation of expertise. Next, we describe the methodology employed in this paper to show how those aggregation problems may be exhibited in engineering decisions, followed by an analysis using an engineering safety

decision as a notional case. At the end, we discuss the implications of the results of the paper for engineering managers, followed by general conclusions.

2. Literature Review

In this section, we review the main expert aggregation techniques found in the literature, together with some of their potential drawbacks and limitations in their use. We start by presenting the discursive dilemma and impossibility result, together with some possible solutions for such a dilemma, and continue to present methods that are based on math to perform the aggregation.

2.1. Discursive Dilemma and the Impossibility Result

The impossibility result [10] has important implications in domains such as political theory, social epistemology, metaphysics [12], economics, logic, and computer science [13]. A common understanding of the impossibility statement is that, when a group of judges hold a rational set of judgements, it is not always possible for them to aggregate their judgements into a collective one in conformity to all possible constraints. When all the judgements of group members are required to be aggregated, it might be logically impossible to rationalize outcomes, which implies that perfect integrity is unattainable. Perfect integrity is the state of constancy and coherence [14]. Coherence is the quality of being logical and consistent, and constancy is the quality of being dependable [15]. Therefore, being unable to achieve perfect integrity implies that logical relationships cannot always be deduced from the aggregated outcome in any aggregation scheme.

An initial step in the task of aggregating judgements is that it should be distinguished from “aggregating people’s sets of credences in respect of certain propositions” [10]. That is, aggregating judgements should be distinguished from the task of belief or opinion elicitation. According to List and Pettit [10], a judgement is “an on or off affair”, meaning that the verdict of an individual concerning a proposition is binary (yes or no, true or false) and should not be mixed with belief of the individual which can be modelled as a probability distribution. Eliciting expert judgement from engineers is, in our experience, common. Examples include, among others, asking experts to assess if the system will meet the performance or not, if the test will be successful or not, if the design will be finished in time or not, or if the material will provide the necessary attenuation or not. Other examples are similar, but asking for a specific value, such as when an analysis may be finished or what power consumption will a system have.

A well-known problem in judgment aggregation is the discursive dilemma [10,13,16]. A simple illustration of the discursive dilemma is shown in Table 1. Let p , q , and r be three logically interconnected propositions, such that $(p \wedge q) \leftrightarrow r$. In a system of three people (1, 2, and 3) making a judgement, each judge expresses his/her judgement on p , q , and r abiding by the judgment rule $(p \wedge q) \leftrightarrow r$. When we observe the aggregated majority of the individual propositions, the group’s collective judgment violates the logic $(p \wedge q) \leftrightarrow r$. The discursive dilemma indicates that a procedure like, for example, systematic majority voting or simple majority voting on each premise cannot guarantee a rational set of collective judgements [10]. However, this type of aggregation is, in our experience, not uncommon in engineering contexts.

Table 1. Illustration of the discursive dilemma.

Judge	p	q	r
Judge 1	True	True	True
Judge 2	True	False	False
Judge 3	False	True	False
Majority	True	True	False

The impossibility result also shows that such problems of deductive cogency and consistency are not only confined to a majority procedure but also persistent in other

aggregation procedures [13,16]. The other aggregation procedures that might be affected by the impossibility results include unanimitarianism (virtue of pertaining to unanimity), minoritarianism and other procedures that have conditional majoritarianism such as two-thirds [10]. This further indicates that if the group follows some behavioral aggregation or even a structured aggregation process [3] of coming to a group consensus, then they might also fall into the problems of the discursive dilemma.

In fact, according to List and Pettit [10], an aggregation method does not exist that jointly satisfies the following set of requirements:

- There must be a universal domain. The aggregation must accept any logical judgement set (personal profile–individual expert’s judgement set).
- The judgements must be collected anonymously. The collective judgement must provide the same result irrespective of the order in which the individual judgement sets are collected.
- The rule of aggregation must be systematic. All individual propositions are treated equally (no special weighting for any individual judgement set).
- The set of judgements must be consistent and deductively closed.

If all these criteria are fulfilled, then the aggregated judgement set can be termed perfect. However, it is impossible for any aggregation method to jointly satisfy all these constraints and yet be collectively rational [10,12]. Such aggregation problems are applicable to most situations that require combining binary evaluations of individual voters.

2.2. Potential Solutions to the Dilemma

The literature states that any aggregation procedure that satisfies the conditions of universal domain, anonymity, and systematicity comes at the cost of collective rationality, consistency, or deductive closure. In addition, there can only be a procedure that roughly approximates adherence to all three principles [10]. However, there are two ways that have been proposed to avoid the paradox: the premise-based procedure (PBP) and the conclusion-based procedure (CBP) [13,17].

In the PBP, the individual judges express their judgement on the set of propositions p and q (that is, the premises in Table 1), the majority of the premise is then aggregated, and a conclusion is drawn from the aggregated premise set. In the PBP, the view taken by the majority on the conclusion will be rejected, thereby ignoring individual responsiveness and adhering to collective rationality [10].

In the CBP, the judges express their final verdict on conclusion r , which abides to the logic $(p \wedge q) \leftrightarrow r$. Then, the majority of the conclusion is aggregated as the final judgement. Thus, in the CBP, significance is given to the group’s conclusion, that is, maintaining individual responsiveness but violating collective rationality [10].

In the literature, there has been considerable support for utilizing the PBP over the CBP. An important result of Hartmann et al. [13] is that, in a voting procedure, the reasons should carry more weight than the conclusion. Mosleh et al. [18] also supported the use of decomposition to utilize experts’ information. It was stated that decomposition is particularly useful when different experts have more information about different aspects of the problem. Raiffa [19] and Armstrong [20] support decomposition, which indicates indirect support of premise-based aggregation. The decomposition method is the technique where the expert is asked to respond to a series of questions on parts of the problem rather than the composite final question. Then, the analyst synthesizes the responses to construct the forecast. While this is a preferred approach in the literature, we have not found any evidence that it proves better results. Note that, regardless of the approach taken, there is still a latent problem in aggregating judgement, since both approaches achieve consistency by ignoring information from the experts.

A potential way to abide by both precepts of individual responsiveness and collective rationality is to practice *modus tollens* instead of *modus ponens* [10]. *Modus tollens* is the propositional logic that states that if a conditional statement (“if p then q ”) is admissible, and the consequent is not true (*not-q*), then the negation of the antecedent (*not-p*) can be inferred.

Modus ponens is the propositional logic that states that if a conditional statement (“if p then q ”) is accepted, and the antecedent (p) holds, then the consequent (q) may be inferred [21]. As in the previous case, there is still a latent problem in aggregating a judgement that is resolved by assuming certain precepts on the experts’ judgement procedure.

2.3. Endorsements for Mathematical Aggregation Techniques over Aggregation of Judgement or Consensus

Mathematical aggregation is the integration of independent opinion assessments into a singular opinion. Opinion here refers to a belief characterization of an uncertain event. The level of complexity in mathematical aggregation procedures varies from simple statistical averaging to approaches based on axiomatic information gathering methods that incorporate the ‘expertise’ of the expert into the aggregation [5,22].

Mosleh et al. [18] classified ‘mathematical aggregation’ as potentially beneficial compared to group consensus methods. In fact, there is conclusive evidence that indicates that mathematical methods for aggregation generally provide better results than the behavioral methods [23–27], yet mathematical aggregation techniques are rarely used in practice [18].

Mosleh et al. [18] presented the concepts of substantive goodness and normative goodness as measures of quality for expert elicitation. Substantive goodness refers to the knowledge of goodness relative to the problem at hand. Normative goodness refers to the expert’s ability to accurately express that knowledge in accordance with the calculus of probabilities. The latter is important to mathematically aggregate opinions. However, choosing appropriate ‘weights’ for the experts, defining proper methods for elicitation, and then combining them might be overly computationally intensive [28] and requires expert guidance for the process itself.

Mathematical aggregation can be generalized into linear and logarithmic opinion pools [5,28,29]. Contextually in belief aggregation, ‘opinion’ has been used extensively and is often referred to as numerical statements that represent the experts’ degrees of belief on a concerned subject [29]. The central idea in belief aggregation is to find a consensus distribution that satisfies a set of reasonable axioms [22,28,30]. In the following sections, linear, logarithmic, and Bayesian information pooling are explained.

Mathematical and statistical aggregation procedures are highly prevalent, yet most of the literature is focused on aggregating probability distributions. That is, when one must aggregate beliefs then a statistical aggregation procedure has been deemed useful. For example, when we want to consider the individual team member’s competence to arrive at the (factually) right conclusion or when the conclusion’s (prior) probability is known, we can rely on Bayesian analysis to aggregate judgements [13,28]. Bayesian analysis is also recommended when experts are exposed to incomplete or even misleading information that shapes their beliefs. Rufo et al. [28] provided a Bayesian procedure to aggregate experts’ information in group decision making. Here, the belief of each expert is elicited as a multivariate prior distribution, followed by a linear or logarithmic combination method to represent a consensus distribution. The choice of the strategy depends on the decision maker.

2.4. Linear and Logarithmic Aggregation (Opinion Pooling)

One common averaging technique for aggregating expertise in the form of opinion is classified as linear opinion pooling. Here, the combined final value that captures the expertise of the group of experts is the weighted linear combination of the probability distribution of the belief of each expert [5,22,29]. The linear opinion pooling was introduced by Stone [31] and can be expressed as

$$\pi(\theta) = \sum_{i=1}^n w_i \pi_i(\theta) \quad (1)$$

where $\pi(\theta)$ is the aggregated probability distribution, θ is the quantity of interest on which the experts express their belief, n represents the number of experts who are involved in the assessment process, $\pi_i(\theta)$ is the i th expert’s individual probability distribution, w_i is the

weight allocated to the i th expert, with weights adding up to 1, and π represents a mass function when θ is discrete and a density function when θ is continuous [22].

The other common form of opinion pooling is logarithmic pooling, where the combined probability distribution can be expressed as

$$\pi_{\log}(\theta) = k \prod_{i=1}^n \pi_i(\theta)^{w_i} \quad (2)$$

where k is the normalizing constant.

In logarithmic opinion pooling, the individual belief distributions are multiplied and renormalized [28]. When the weights are equal to $(1/n)$, then $\pi_{\log}(\theta)$ is proportional to the geometric mean of the individual distributions. A generalization of the linear and logarithmic opinion pools is presented in [32].

Linear and logarithmic opinion pooling have been used in a wide variety of domains such as medical consultation [33], marketing, banking, weather forecasting [22,34], and for candidate selection of football games [35]. In most of these applications, the weights of the experts were determined using empirical studies and were usually based on the subjective trust or confidence that the decision maker has on the reliability of the experts [5]. Linear and logarithmic pooling often lead to distinctively different distributions [28]. However, there seems to be no exceptional advantage of one method over the other [36]. One seeming advantage of logarithmic pooling over linear pooling is that, when combining density functions, the results are always unimodal, whereas linear pooling of density functions can produce combined distributions that are multimodal in nature and may cause a bias for the analyst [5]. Unimodality can be interpreted here as a more accurate representation of the collective beliefs of the experts, as the logarithmic effects dampen strong differences in the assessments.

Other probability combination techniques based on other statistical methods, such as frequency theory, have also been studied for opinion pooling [29]. However, for circumstances such as catastrophic disasters, common in engineering decisions, where the interpretation of frequencies by the expert is stretched to the limits of plausibility, they present greater difficulties to accurately represent the expert beliefs quantitatively. Furthermore, it has been established that frequency theory fails when data are sparse, unavailable, or subjected to non-sampling errors, which are common scenarios in engineering. Such problems gave endorsements to the development of Bayesian aggregation methods [29].

2.5. Advocacy for Bayesian Methods

There is extensive research in using a Bayesian aggregation scheme to support risk analysis that requires a group of experts to provide information to a decision maker [29,37,38]. When a prior probability distribution is available for the parameter of interest, then a Bayesian technique can be applied to update the prior distribution with a combined opinion pool [28]. The main advantage of Bayesian methods is that they allow for the incorporation of the ‘expertise’ level of the expert and dependencies among the experts into the aggregation model [22]. With n experts, providing information regarding a parameter of interest θ and the probability distribution of θ is known in prior $\pi(\theta)$, then the analyst/decision maker can make use of Bayes’ Theorem to update $\pi(\theta)$ [22] using the following relationship:

$$\pi^* = \pi(\theta|i_1, i_2, \dots, i_n) \propto \frac{\pi(\theta)L(i_1, i_2, \dots, i_n|\theta)}{\pi(i_1, i_2, \dots, i_n)} \quad (3)$$

where L is the likelihood function associated with the experts’ information.

However, Bayesian-based methods have been shown to be difficult to apply in practice [22] due to their computational complexity [5] and the difficulty of estimating the likelihood function L , since it must account for the prediction accuracy and bias of the individual expert.

3. Application of Expert Aggregation to Safety Decisions

In this section, we apply some of the behavioral and mathematical aggregation schemes discussed in the previous section to notional cases that involve a panel of engineers assessing the criticality of safety hazards. Safety is used because safety-related decisions are common to most engineering endeavors. The different aggregation schemes are used to show the different effects that such schemes have on the resulting decisions. The application of the chosen case and the different methods are representative of their application in practice as per the experience of the authors.

3.1. Case Description

We used the SRMPs of the FAA and the U.S. Navy [7,8] as a reference for the safety assessment process. Specifically, we explored cases in which the experts were tasked to collectively categorize the criticality of a safety hazard as a function of the severity of its consequences and its likelihood of occurrence (in this case through a risk matrix, as shown in Figure 1, or a version of it). A conceptual argument was presented to verify if those solutions were applicable to the safety risk assessment. We studied the conditions under which the different aggregation methods preserved or failed to preserve the premise and consistency.

Severity Likelihood	Minimal 5	Minor 4	Major 3	Hazardous 2	Catastrophic 1
Frequent A	[Green]	[Yellow]	[Red]	[Red]	[Red]
Probable B	[Green]	[Yellow]	[Yellow]	[Red]	[Red]
Remote C	[Green]	[Green]	[Yellow]	[Yellow]	[Red]
Extremely Remote D	[Green]	[Green]	[Green]	[Yellow]	[Red] [Yellow]
Extremely Improbable E	[Green]	[Green]	[Green]	[Green]	[Yellow]

High Risk [Red]

Medium Risk [Yellow]

Low Risk [Green]

* High Risk with Single Point and/or Common Cause Failures

Figure 1. Small aircraft/rotorcraft safety risk matrix (source: [7] Appendix C, Figure C-2).

The following aggregation methods were studied:

- Judgement aggregation:
 - a. CBP and PBP;
 - b. Modus tollens vs. modus ponens;
 - c. Techniques to evade the dilemma;
 - i. Convergence;
 - ii. Authoritarian;
 - iii. Priority;
 - iv. Special support.

- Mathematical aggregation:
 - a. Arithmetic and geometric aggregation of severity;
 - b. Belief aggregation for likelihood of hazard occurrence;
 - c. Belief aggregation using linear pooling;
 - d. Belief aggregation using logarithmic pooling.

For the assessments of likelihood, we used the definitions provided by the FAA: likelihood is “the estimated probability or frequency, in quantitative or qualitative terms, of a hazard’s effect or outcome” [39]. Table 2 provides the likelihood definitions followed by the FAA based on which the experts assert their assessments for determining the safety risk of a given hazard.

Table 2. Likelihood definitions (source: Appendix C Table C2. [39]).

	Qualitative	Quantitative Time/Calendar-Based Occurrences Domain-Wide/System-Wide
Frequent A	Expected to occur routinely	Expected to occur more than 100 times per year (or more than approximately 10 times a month)
Probable B	Expected to occur often	Expected to occur between 10 and 100 times per year (or more than approximately 1–10 times a month)
Remote C	Expected to occur infrequently	Expected to occur one time every 1 month to 1 year
Extremely Remote D	Expected to occur rarely	Expected to occur one time every 1 to 10 years
Extremely Improbable E	Unlikely to occur, but not impossible	Expected to occur less than one time every 10 years

All data used in the studies are synthetic yet reasonable.

3.2. Judgement Aggregation

A notional case study involving a panel of three experts was presented to demonstrate the possibility of the discursive dilemma (a problem giving rise to premise and judgement inconsistencies) in a safety decision-making scenario. Their judgements on severity and likelihood of an assumed hazard were synthetically established and the safety risk was determined through a logical relationship between safety risk, severity, and likelihood. The logical relationship was derived with reference to the safety risk matrix presented in [7]. The solutions to overcome the dilemma presented in earlier were applied to this case and a ‘*what if*’ scenario analysis analogous to the practices in project management [40] was performed. The critical impacts of the potential ‘solution’ methodology were discussed.

3.2.1. Base Case

Consider a safety risk assessment required to be performed by a team of three engineers. The judgement of all three engineers is assumed to be equally trustable and relevant for the decision. Using the risk matrix in Figure 1, the hazard is currently considered *High Risk* because its severity is judged as level 3 (Major) and its likelihood of occurrence is judged as A (Frequent). The experts are asked to judge such assessment.

To avoid issues of internal consistency, we assumed without loss of generality that the judgement set of each engineer (that is, the severity, likelihood, and safety risk) satisfied the three conditions of completeness, consistency, and deductive closure [10]. A personal judgement set is said to be complete, consistent, and deductively closed if, for all propositions proposed by the expert, the final judgement is available in the universal domain set (high, medium, and low risk), it conforms to a logical deduction, and the logic remains true for the combinations of the propositions. In this context, the goal for aggregating the

engineers' expertise is to find an aggregation technique that conforms to the minimum requirements of completeness, consistency, and deductive cogency [10].

The engineers' individual judgements, prescribed by the SRMPs coded in Figure 1, follow the logic given below (note that the symbol ' $\wedge$ ' represents the AND operation):

$$(\text{Severity (Major 3)} \wedge \text{Likelihood (Frequent A)}) \rightarrow \text{Safety Risk (High Risk)}$$

Assume now, without loss of generality, that the three engineers provide assessments as given in Table 3, and that the decision maker uses majority voting to arrive at a conclusion (shown in the last row of the table). In such a case, as shown in Table 3, the requirements of *deductive cogency* and *consistency* are violated. The majority asserts that the severity of the hazard is *Major*, and that the likelihood of occurrence is *Frequent*. However, the majority rules out the proposition that the safety risk of the hazard is *High*. Therefore, the aggregated assessment of the engineers is illogical. We discuss next the implications of aggregating only the conclusions or aggregating only the conditions as potential mechanisms to avoid illogical reasoning.

Table 3. Discursive dilemma in safety risk assessment.

SME	Severity <i>Major (3)</i>	Likelihood <i>Frequent (A)</i>	Safety Risk <i>High Risk</i>
1	Yes	Yes	Yes
2	Yes	No	No
3	No	Yes	No
Majority	Yes	Yes	No

3.2.2. Evaluation of Potential Solutions to the Discursive Dilemma

Practicing CBP and PBP

If the CBP is being practiced, then the individual engineers perform their personal assessment of severity and likelihood to assess the criticality of the hazard, but only communicate to the decision maker their assessment of criticality, not those of severity or likelihood. The decision maker then aggregates their criticality assessment (i.e., the engineers' conclusions) to determine a criticality level for the hazard that reflects the engineers' assessments. The result of the CBP with respect to the premises of the engineers presented in Table 3 is shown in Table 4. The result shows that the majority would decide that the hazard is *NOT* High Risk.

Table 4. Aggregated judgement for CBP.

SME	Safety Risk <i>High Risk</i>
1	Yes
2	No
3	No
Majority	No

When the PBP is practiced, the group's assessments of the severity and likelihood are first aggregated and then a conclusion is drawn from the aggregation. In other words, engineers are asked in this case to assess the causes (i.e., hazard severity and likelihood), not the conclusions (i.e., hazard criticality). Table 5 shows the application of the PBP for the case shown initially in Table 3. In this case, the (aggregated) hazard is assessed as *High Risk* based on the logical relationship shown earlier.

Table 5. Aggregated judgement for PBP.

SME	Severity <i>Major (3)</i>	Likelihood <i>Frequent (A)</i>
1	Yes	Yes
2	Yes	No
3	No	Yes
Majority	Yes	Yes

From this, it can be seen that the same experts having the same judgement can potentially end up with different assessments depending on the questions asked. This indicates inconsistency and retraces back to the problems in expert elicitation [41]. A normative aggregation method, however, should yield the same assessment, regardless of the method of elicitation, and should abide by the logic used to derive the judgement.

In these cases, Hartmann et al. [13] suggest that, in a voting procedure, the reasons should carry more weight than the conclusion. This implies that collective rationality should be maintained and carries more significance. Therefore, we can choose to forego adherence to individual responsiveness. If this reasoning is adopted to make safety decisions in engineering, following the PBP leads to taking the *safer* route, since the relationship between the conclusion and the premises is essentially a logical *AND* operation. In a worst-case scenario, hence, the PBP leads to judge a safety risk higher than the individual engineers intended.

Practicing *modus tollens* instead of *modus ponens*

Although practicing *modus tollens* could be a possible technique to avoid the dilemma, it could have negative implications in safety risk assessments. If *modus tollens* is practiced, then engineers are asked to assess either the severity or the likelihood, and their opinion on the criticality of the hazard. Then, the unquestioned variable (either the severity or the likelihood) is inferred from the relation. This method maintains deductive cogency and collective reasoning, evading the dilemma.

Following the earlier example, consider now that the engineers are asked about their agreement with the criticality assessment of hazard. If their answer is affirmative, then the severity of the hazard is *Major (3)* and the frequency of the hazard is *Frequent (A)*. If one parameter is fixed, then the *modus tollens* method can be used to identify an inconsistency in the judgement of an SME. Consider the judgement set of SME 3 in Figure 2. The SME thinks that the safety risk posed by the hazard is severe but does not seem to accept that the frequency of the hazard is Frequent. Such problems are attributed to inconsistencies in expert elicitation and behavioral aggregation, where the judges may violate principles of deductive cogency without realizing its implications.

Strategies that yield collectivized reason [10]

SME	Severity <i>Major (3)</i>		Likelihood <i>Frequent (A)</i>		Safety Risk <i>High Risk</i>
1	Yes	$\wedge$	Yes	$\leftrightarrow$	Yes
2	No	$\wedge$	Yes	$\leftrightarrow$	No
3	Yes	$\wedge$	No?	$\leftrightarrow$	Yes

Figure 2. Modus tollens.

Using a convergence strategy. By practicing interpersonal deliberations or by other methods, if the views of the engineers are made to converge, then the conclusion of the collective majority will be complete, consistent, and deductively closed. By doing so, we relax the universal domain. This suggests adopting a behavioral technique where the group comes to a consensus on the engineering assessment. However, while this strategy can be useful to avoid the paradox, it may do so at the expense of becoming unreliable. When interactions occur among the engineers, it might be difficult to distinguish

if convergence occurs because the knowledge base of the engineers has unified or because the group is biased towards the same opinion. It is also difficult to acknowledge group-level consequences that emerge as a result of individual interaction systems [42]. Finally, the group interactions could also give rise to a number of cognitive biases [43] that might go undetected.

Using an authoritative strategy. In the example, where majority voting was used, each engineer was considered to have equal weight and all their assessments were treated as equally important. However, if modifying the group structure is permitted so that only the input of some members is taken into consideration, then it is easy to avoid the paradox. This is achieved by relaxing the rule of anonymity. This, however, is still subject to the problems of assigning proper weights to the experts. Furthermore, considering the assessments of only a limited number of engineers will place into question why the others were included in the assessment in the first place and ignored later. Ignoring those engineers may lead to selection biases [44], where only information that the decision maker wants to *hear* is used to make the decision. Essentially, using an authoritative strategy may easily result in not utilizing the expertise of the engineers involved in the decision.

Using a priority strategy. In this strategy, the team must decide on a set of propositions to be given higher priority over the others. The team's decision process of the other set of propositions differs from the prioritized set and is determined on the overall decision of the prioritized set of positions. If a conclusive judgement set is contradictory and/or does not belong to the prioritized judgement set, then the team eliminates the judgement set that has a lower priority, thus eliminating its impact on the final judgment. By following a priority strategy, the rule of systematicity is relaxed. The implication of relaxing the rule of systematicity can be that the logic used to arrive at the aggregated judgement might be inconsistent. Although the dilemma might be evaded, the rationale behind the aggregated judgement might be incorrect.

Using a special-support strategy. In this strategy, a proposition set proposed by each expert must be endorsed or receive *special support* by the majority of the other experts. This strategy can be thought of as a *majority of the majorities*. Each expert presents his/her judgment set to the remainder of the group and asks for endorsements. The judgement set that has the highest number of endorsements or amount of support from the *super-majority* is concluded as the final collective judgement. By using this strategy, the rule of completeness is relaxed. However, the collective assessment may still be flawed if the engineer receiving the *special support* does not have a logical judgement. Furthermore, this strategy also suffers from a potential bias of group thinking.

A summary of the implications of using each of these four strategies in safety risk assessment is given in Table 6.

Table 6. Aggregation strategies that yield *Collective Reason*.

Strategy	Implications
Convergence	Unreliable method: gives rise to potential group biases that might go undetected.
Authoritative	Assigning weights or determining the authoritarian might lead to <i>selection bias</i> .
Priority	Logic between premise and conclusion might become deterred.
Special support	Rationale behind the safety risk assessment might be illogical; leads to potential group bias.

3.2.3. Implications of the Dilemma

The FAA and the U.S. Navy prescribe that, once a safety risk has been assessed and documented, the decision process is solely based on the documented value of the safety risk [7,8]. If the hazard has been determined as *High Risk*, the decision process is assigned to higher authoritarian members of the organization. If the safety risk has been assessed as *Medium* or *Low*, then the decision authority or advisors responsible for providing instruction

for the mitigation and control activities of the hazard are members of lower authoritative ranks. Hence, the dilemma poses a risk that the rightful authorities might not be called to act on the hazard that has been incorrectly judged as *Medium*, *Low*, or *High*.

Furthermore, when undesirable outcomes emerge as a consequence of the decision based on the safety risk assessment, the aggregation procedure might not be questioned but rather the SMEs judgement (prediction accuracy) might be deemed unreliable. This results in the ability to improve decision-making processes.

3.3. Mathematical Aggregation

3.3.1. Arithmetic and Geometric Averaging of Severity Rankings

Consider a notional case where an engineering manager requests five engineers to assess the severity of three hazards. Assume that the engineering manager aggregates the assessment using arithmetic (AM) and geometric (GM) averaging, as shown in Figure 3.

Expert	Hazard X	Hazard Y	Hazard Z
Expert 1	4	1	5
Expert 2	3	1	3
Expert 3	1	4	1
Expert 4	3	2	2
Expert 5	5	5	2
AM	3	3	3
GM	3	2	2

Figure 3. Severity assessments for hazards X, Y and Z by five experts.

First, the results show that the choice of averaging technique yields different results for some of the hazards. This is due to the calculation procedure behind the averaging method. The AM is the sum of the assessments divided by the number of assessments and the GM is the n th root of the product of the assessments, where n is the number of assessments. In geometric averaging, the effect of ‘outliers’ is dampened. However, since the possible set of assessments here lies only in the range of 1 to 5, the characteristic of dampening outliers by the GM might not make a critical impact.

In aggregating severity rankings, when the assessments of the experts vary between extreme points, then applying either the AM or the GM discounts the assessment of the expert. For instance, in the assessments of Hazard X and Hazard Z (ref. Figure 3), Expert 3 has assessed the severity of that hazard to be catastrophic. However, when the assessments of all the experts were averaged, the aggregated assessment indicated that the hazard was not as ‘severe’ as anticipated by Expert 3. In a situation where Expert 3 has the highest prediction accuracy, then the averaged aggregated assessment has discounted the input of Expert 3 and has yielded a lower severity level, thus leading to an improper safety risk assessment and thereby to inadequate mitigation and control strategies required for the hazard.

Furthermore, it is worth noting that assessing rankings is a form of judgement. Therefore, this kind of ranking inherits the problems associated with judgement presented earlier, regardless of the sophistication of the mathematical approach used to aggregate the judgments.

Finally, there are mathematical issues when averaging rank-ordered severity assessments where the scale between the values is not linear. The notion of $(Low + High)/2 = Medium$ does not hold when the scale is not linear, which is generally the case, and particularly true for the definitions from the literature used in this paper.

3.3.2. Arithmetic and Geometric Averaging of Likelihood Rankings

Consider a similar case to the previous one, but this time the engineering manager asks their team of engineers to assess the likelihood of different options, and the judgements shown in Figure 4.

Expert	Hazard X	Hazard Y	Hazard Z
Expert 1	4	1	5
Expert 2	3	1	3
Expert 3	1	4	1
Expert 4	3	2	2
Expert 5	5	5	2
AM	3	3	3
GM	3	2	2

Figure 4. Likelihood assessments for Hazards X, Y and Z by five experts.

First, the problems indicated earlier when using the AM and the GM to aggregate severity assessments also apply to aggregating likelihood assessments. Furthermore, there are some mathematical effects due to the underlying mathematical structures of beliefs and likelihoods that can easily introduce significant errors in the assessment of likelihoods.

For example, when two experts assert the likelihood of a hazard being *Probable*, their probability (belief) distribution could be quite different from each other's. (Note that objective probabilities do not exist, so it is sensible that each expert will have their own belief distribution [2].) For instance, both belief distributions could follow a Poisson process yet can have different mean rates of occurrence. Similarly, both engineers could have belief distributions that are different in shape but with the same mean rates. Or even, different engineers may use a different target to define the meaning of each likelihood category (e.g., different engineers assign different confidence targets to the occurrence rate of the event of a given likelihood category). Therefore, even though their assessment may be identical in the risk table, their internal interpretation is different.

3.3.3. Belief Aggregation Using Linear Pooling

Consider the same example but this time, instead of assessments, the experts' beliefs on the likelihood of the hazards are elicited as belief distributions. For simplicity and without lack of generality, let us assume these are defined as dichotomic, assessing the probability of a hazard existing to be 36%, 21%, and 45%, respectively.

With linear pooling, the beliefs of the three experts are aggregated using the following formula:

$$\pi_{linear}(k) = \sum_{i=1}^n w_i \pi_i(k) \quad (4)$$

where $n = 3$ and $w_1 + w_2 + w_3 = 1$.

Let us assume, as a start, that the engineering manager trusts each expert equally, so he/she decides to assign $w_1 = w_2 = w_3$. This leads to an aggregated probability assessment for the hazard of 34%.

Mathematically, linear pooling can be considered to be well constructed. However, its use must meet several conditions for it to be meaningful. In particular, it is key to understand both the purpose of the aggregation of beliefs and the way in which each expert's individual belief is formed. Let us start with one scenario that showcases these aspects.

Assume that the three experts have the same education, and their backgrounds only differ in the projects they have worked on. Experts 1 and 2 worked together on the same project (Project A), where the hazard under assessment did not occur. Expert 3, on the contrary, has worked on three projects (Projects B, C, and D), and the hazard occurred in two of them. Given that experts 1 and 2 are basing their beliefs on the "same" information, the engineering manager is effectively overcounting the particular experiences of Project

A over Projects B, C, and D, which results in skewing the belief aggregation (even if the individual beliefs of the experts are accurate).

Linear pooling makes use of weights to calibrate or adjust aggregation to take these aspects into account. Because, while probabilities are subjective, they must not be arbitrary, meaning it is important to understand how engineers form their beliefs, so that the engineering manager can account for individual biases and for biases injected as part of the process of aggregation. Individual biases are discussed at length in the literature (e.g., [2]), and will not be addressed in this paper. For biases in the aggregation of beliefs, we focused on the alignment of the purpose of the aggregation and the bases to form experts' beliefs.

An engineering manager may desire to aggregate beliefs for two reasons. First, to calibrate the assessments of the experts. This means that instead of relying on the assessment of one expert, aggregating the beliefs of several experts can be used to filter the errors in their assessments. Here, error does not refer to a delta between the expert's assessment and an objective reality, but the delta between the elicited, tangible belief and their real, inner belief. Second, to comprehensively account for disparate experiences. This means that the engineering manager attempts to avoid overcounting or skewing the belief assessment on a particular set of experiences, reaching instead a wide variety of *data points*.

When attempting to calibrate experts' assessments, it is then essential that experts use the same set of information when their beliefs are elicited. Otherwise, one gets into the problems shown in the previous example. It is the engineering manager's responsibility then to guarantee that experts have access to the same information set. For example, an engineering manager could ask all experts to explain how they arrived at their belief (without sharing what their belief is, to avoid injecting individual biases), so that the information is levelled across all experts. Without doing so, the validity of using linear pooling to calibrate experts' assessments is jeopardized.

On the contrary, when attempting to account for different experiences, experts should use non-overlapping information. That is, the information set of each expert should be mutually exclusive, in order to avoid the overcounting of some experiences. For example, consider three experts providing their expertise to a decision. Two of the experts have had the same experience on the same project. Aggregating their expertise would be akin to double counting what happened in just that one project. Avoiding overcounting, in practice, is virtually impossible given that experts will likely share some educational or professional baseline. As in the previous case, it is the engineering manager's responsibility to assess the baseline information that each expert is using to assess if the aggregation is valid or not. Again, in this case, information sharing may be a valuable technique to at the same time be comprehensive in the elicitation and calibrate it.

3.3.4. Belief Aggregation Using Logarithmic Pooling

Logarithmic pooling further emphasizes differences in which expert's opinion to account for. Therefore, the same discussion as for linear pooling is applicable here.

4. Conclusions

We have shown that common techniques to aggregate expertise from a team of engineers lack internal consistency and lead to the inadequate use and aggregation of engineering expertise. Particularly in safety risk assessment, the plausibility of the discursive dilemma and its implications were demonstrated. We showed that, if a judgement aggregation must take place, then a '*better*' process to evade the dilemma would be the PBP (premise-based procedure). However, all solutions explored had drawbacks, since all solutions operate by ignoring information from experts, which includes either ignoring individual responsiveness or violating collective rationality.

In general, we showed how using the current definitions of severity, which are abstract and qualitative, and using them to perform a mathematical aggregation of the severity judgments leads to problems associated with the proper use of each expert. Similarly, methods to aggregate likelihood assessments that are based on judgement led to an inaccurate

aggregation. Even those that are based on mathematical formalization with a sound base may become inadequate if the underlying conditions for the aggregation are not made explicit and validated. Such inconsistencies lead to the mistreatment of expert knowledge, by ignoring experts' knowledge, emphasizing certain experts over other experts without a good reason to do so, or double counting single experiences.

Overall, this paper has shown the importance of understanding how the aggregation of expertise is performed. An engineering manager should not simply rely on consulting a group of experts and try to make a sort of voting or simple averaging. Instead, information must be elicited individually, and the purpose of the aggregation aligned with the experts that are consulted and the information that is elicited.

We suggest future research to empirically evaluate the use and impact of expert aggregation techniques in practice, that is, in real applications, including more granularity of methods, particularly those that rely on mathematics to perform the aggregation.

Author Contributions: Conceptualization, C.S. and A.S.; methodology, C.S. and A.S.; formal analysis, C.S.; writing—original draft preparation, C.S.; writing—review and editing, H.K. and A.S.; supervision, H.K. and A.S.; project administration, A.S.; funding acquisition, A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by NAVAIR Contract # N00421-16-2-001.

Data Availability Statement: No new data were created in this research.

Acknowledgments: The authors gratefully acknowledge Susan J. DeGuzman, the USN/USMC Airworthiness and CYBER-SAFE Office, and the Naval Air Systems Command.

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

Nomenclature

The following nomenclature are used in this manuscript:

Acronym	Description
AM	Arithmetic Averaging
CBP	Conclusion Based Procedure
FAA	Federal Aviation Administration
GM	Geometric Averaging
PBP	Premise-Based Procedure
SME	Subject Matter Expert
SRMP	Safety Risk Management Policy

References

1. Vismari, L.F.; Junior, J.B.C. An Absolute-Relative Risk Assessment Methodology Approach to Current Safety Critical Systems and its Application to the ADS-B based Air Traffic Control System. In Proceedings of the 2008 Symposium on Reliable Distributed Systems, Naples, Italy, 6–8 October 2008; pp. 95–104.
2. Howard, R.A.; Abbas, A.E. *Foundations of Decision Analysis*, 1st ed.; Pearson: London, UK, 2015.
3. van Steen, J.F.J. A perspective on structured expert judgment. *J. Hazard. Mater.* **1992**, *29*, 365–385. [\[CrossRef\]](#)
4. Mak, B.; Bui, T.; Blanning, R. Aggregating and updating experts' knowledge: An experimental evaluation of five classification techniques. *Expert Syst. Appl.* **1996**, *10*, 233–241. [\[CrossRef\]](#)
5. Chytka, T.M. Development of an Aggregation Methodology for Risk Analysis in Aerospace Conceptual Vehicle Design. In *Engineering Management*; Old Dominion University: Norfolk, VA, USA, 2003.
6. Leung, K.; Verga, S. Expert judgement in risk assessment. *Def. RD Can. Cent. Oper. Res. Anal.* **2007**.
7. *FAA-Order8040.4B*; Safety Risk Management Policy. Federal Aviation Administration: Washington, DC, USA, 2017.
8. *Navy-OPNAVINST3500.39C*; Operational Risk Management. 2000 Navy Pentagon: Washington, DC, USA, 2010; pp. 20350–22000.
9. Rumsfeld, D.H. AVIATION SAFETY FAA and DOD Response to Similar Safety Concerns. *Aviat. Saf.* **2002**.
10. List, C.; Pettit, P. Aggregating sets of judgments: An impossibility result. *Econ. Philos.* **2002**, *18*, 89–110. [\[CrossRef\]](#)
11. Phillips, L.D.; Phillips, M.C. Facilitated work groups: Theory and practice. *J. Oper. Res. Soc.* **1993**, *44*, 533–549. [\[CrossRef\]](#)
12. Weintraub, R. A solution to the discursive dilemma. *Philos. Stud.* **2011**, *152*, 181–188. [\[CrossRef\]](#)
13. Hartmann, S.; Pigozzi, G.; Sprenger, J. Reliable Methods of Judgement Aggregation. *J. Log. Comput.* **2010**, *20*, 603–617. [\[CrossRef\]](#)
14. Kornhauser, L.A.; Sager, L.G. The many as one: Integrity and group choice in paradoxical cases. *Philos. Public Aff.* **2004**, *32*, 249–276. [\[CrossRef\]](#)

15. Stevenson, A. *Oxford Dictionary of English*; Oxford University Press: Cary, NC, USA, 2010.
16. List, C.; Puppe, C. *Judgment Aggregation: A Survey*; Oxford University Press: Oxford, UK, 2009.
17. Bovens, L.; Rabinowicz, W. Democratic answers to complex questions—an epistemic perspective. *Synthese* **2006**, *150*, 131–153. [\[CrossRef\]](#)
18. Mosleh, A.; Bier, V.M.; Apostolakis, G. A critique of current practice for the use of expert opinions in probabilistic risk assessment. *Reliab. Eng. Syst. Saf.* **1988**, *20*, 63–85. [\[CrossRef\]](#)
19. Raiffa, H. Decision analysis: Introductory lectures on choices under uncertainty. *J. Financ.* **1968**, *24*, 1000–1003. [\[CrossRef\]](#)
20. Armstrong, J.S. *Long-Range Forecasting*; Wiley New York ETC: New York, NY, USA, 1985.
21. Yalcin, S. A counterexample to modus tollens. *J. Philos. Log.* **2012**, *41*, 1001–1024. [\[CrossRef\]](#)
22. Clemen, R.T.; Winkler, R.L. Combining probability distributions from experts in risk analysis. *Risk Anal.* **1999**, *19*, 187–203. [\[CrossRef\]](#)
23. Von Holstein, C.-A.S.S. Probabilistic forecasting: An experiment related to the stock market. *Organ. Behav. Hum. Perform.* **1972**, *8*, 139–158. [\[CrossRef\]](#)
24. Brown, T.A. *An Experiment in Probabilistic Forecasting*; Rand Corporation: Santa Monica, CA, USA, 1973.
25. Dalkey, N.C. *Group Decision Theory*; California University Los Angeles School of Engineering and Applied Science: Los Angeles, CA, USA, 1977.
26. Winkler, R.L.; Makridakis, S. The combination of forecasts. *J. R. Stat. Soc. Ser. A* **1983**, *146*, 150–157. [\[CrossRef\]](#)
27. Makridakis, S.; Winkler, R.L. Averages of forecasts: Some empirical results. *Manag. Sci.* **1983**, *29*, 987–996. [\[CrossRef\]](#)
28. Rufo, M.J.; Pérez, C.J.; Martín, J. A Bayesian approach to aggregate experts' initial information. *Electron. J. Stat.* **2012**, *6*, 2362–2382. [\[CrossRef\]](#)
29. Genest, C.; Zidek, J.V. Combining probability distributions: A critique and an annotated bibliography. *Stat. Sci.* **1986**, *1*, 114–135.
30. O'Hagan, A.; Buck, C.E.; Daneshkhah, A.; Eiser, J.R.; Garthwaite, P.H.; Jenkinson, D.J.; Oakley, J.E.; Rakow, T. *Uncertain Judgements: Eliciting Experts' Probabilities*; John Wiley & Sons: Hoboken, NJ, USA, 2006.
31. Stone, M. The opinion pool. *Ann. Math. Stat.* **1961**, *32*, 1339–1342. [\[CrossRef\]](#)
32. Cooke, R. *Experts in Uncertainty: Opinion and Subjective Probability in Science*; Oxford University Press on Demand: Oxford, UK, 1991.
33. Winkler, R.L.; Poses, R.M. Evaluating and combining physicians' probabilities of survival in an intensive care unit. *Manag. Sci.* **1993**, *39*, 1526–1543. [\[CrossRef\]](#)
34. Stael von Holstein, C.-A.S. *Assessment and Evaluation of Subjective Probability Distributions*; The Economic Research Institute at the Stockholm School of Economics: Stockholm, Sweden, 1970.
35. Hurley, W.; Lior, D. Combining expert judgment: On the performance of trimmed mean vote aggregation procedures in the presence of strategic voting. *Eur. J. Oper. Res.* **2002**, *140*, 142–147. [\[CrossRef\]](#)
36. Rantilla, A.K.; Budescu, D.V. Aggregation of expert opinions. In Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences, HICSS-32. Abstracts and CD-ROM of Full Papers, Maui, HI, USA, 5–8 January 1999; p. 11.
37. French, S. *Group Consensus Probability Distributions: A Critical Survey*; University of Manchester, Department of Decision Theory: Manchester, UK, 1983.
38. Lindley, D.V. Reconciliation of discrete probability distributions. *Bayesian Stat.* **1985**, *2*, 375–390.
39. *FAA-Order8040.6; Unmanned Aircraft Systems Safety Risk Management Policy*. Federal Aviation Administration: Washington, DC, USA, 2019.
40. Snyder, C.S. *A Guide to the Project Management Body of Knowledge: PMBOK (®) Guide*; Project Management Institute: Newtown Square, PA, USA, 2014; pp. 128–129.
41. Goodwin, P.; Wright, G. *Decision Analysis for Management Judgment*; Wiley: Hoboken, NJ, USA, 2014.
42. Davis, J.H. Some compelling intuitions about group consensus decisions, theoretical and empirical research, and interpersonal aggregation phenomena: Selected examples 1950–1990. *Organ. Behav. Hum. Decis. Process.* **1992**, *52*, 3–38. [\[CrossRef\]](#)
43. Das, T.; Teng, B.S. Cognitive biases and strategic decision processes: An integrative perspective. *J. Manag. Stud.* **1999**, *36*, 757–778. [\[CrossRef\]](#)
44. Norton, M.I.; Sommers, S.R.; Brauner, S. Bias in jury selection: Justifying prohibited peremptory challenges. *J. Behav. Decis. Mak.* **2007**, *20*, 467–479. [\[CrossRef\]](#)

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