



HAL
open science

Facts versus Interpretations in Intelligence: A Descriptive Taxonomy for Information Evaluation

Benjamin Icard

► **To cite this version:**

Benjamin Icard. Facts versus Interpretations in Intelligence: A Descriptive Taxonomy for Information Evaluation. *Intellectica - La revue de l'Association pour la Recherche sur les sciences de la Cognition (ARCo)*, 2023, *Cognition and Intelligence*, 78 (1), pp.89-105. 10.31234/osf.io/q8avs . hal-04442314

HAL Id: hal-04442314

<https://hal.science/hal-04442314v1>

Submitted on 6 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Facts versus Interpretations in Intelligence: *A Descriptive Taxonomy for Information Evaluation*

Published in *Intellectica*, volume 78 (*Special issue on Cognition & Intelligence*)*

Benjamin Icard

Institut Jean-Nicod (CNRS, ENS-PSL, EHESS)

ABSTRACT

Traditionally, intelligence officers use an alphanumeric scale known as the Admiralty System to evaluate informational messages by rating the credibility of their content and the reliability of their source [e.g. [NATO AJP-2.1, 2016](#)]. Amongst other duties, they are expected to clearly distinguish objective facts from subjective interpretations during this evaluation [e.g. [NATO STANAG-2511, 2003](#)]. That being said, various experimental results show that officers are unable to properly fulfill this methodological duty [e.g. [Baker et al., 1968](#), [Kelly and Peterson, 1971](#), [Johnson, 1973](#)]. Our explanation is that the extant scale, which is evaluative by nature, does not allow them to endorse a more objective, that is to say descriptive, perspective on information. In this article, we aim to help enforce the *facts versus interpretations recommendation* in the intelligence domain. By extracting the descriptive dimensions that underlie the scale, and by grouping them by linguistic directionality [e.g. [Teigen and Brun, 1995](#), [Mandel et al., 2022](#)], we introduce a taxonomy to categorize intelligence messages more objectively. This taxonomy is fine-grained: it integrates messages which are informative or deceptive in the classical sense (e.g. *misinformation*, *lying*), but also more borderline messages, such as *omissions* and *half-truths*, which rely on the use of linguistic vagueness [following [Égré and Icard, 2018](#), [Icard et al., 2022](#)]. By putting descriptive lenses on information evaluation, we seek to provide new categories to help officers make more acute evaluations of information.

KEYWORDS: Intelligence, Information Evaluation, Admiralty Scale, Descriptive vs. Evaluative, Vagueness, Taxonomy.

RÉSUMÉ

Dans le domaine du renseignement, les officiers en charge d'évaluer l'information utilisent une échelle alphanumérique traditionnellement appelée "Admiralty System", pour noter la crédibilité du contenu de l'information, d'une part, et la fiabilité de sa source, d'autre part [e.g. [NATO AJP-2.1, 2016](#)]. Au cours de cette évaluation, il leur est notamment demandé de bien distinguer les faits objectifs des interprétations subjectives qu'ils en font [e.g. [NATO STANAG-2511, 2003](#)]. Cela étant, divers résultats expérimentaux montrent que les officiers échouent à respecter cette recommandation de nature méthodologique [e.g. [Baker et al., 1968](#), [Kelly and Peterson, 1971](#), [Johnson, 1973](#)]. Une hypothèse assez intuitive est que l'échelle existante, évaluative par nature, ne leur permet pas d'adopter une perspective plus objective, disons descriptive, sur l'information qu'ils reçoivent. Le but de cet article est de faciliter la *distinction entre faits et interprétations* lors de l'évaluation du renseignement. En isolant les dimensions descriptives qui sous-tendent l'échelle existante, puis en les groupant en fonction de leur directionnalité [e.g. [Teigen and Brun, 1995](#), [Mandel et al., 2022](#)], nous proposons une taxonomie de l'information pour catégoriser les messages de renseignement plus objectivement. Cette taxonomie est fine: elle intègre les messages visant à informer ou à tromper de façon classique (e.g. *mésinformation*, *mensonge*), mais aussi les messages relevant de cas plus limites, comme l'*omission* et les *demi-vérités*, qui sont fondés sur l'usage du vague linguistique [à partir de [Égré and Icard, 2018](#), [Icard et al., 2022](#)]. En donnant une vision plus descriptive de la procédure d'évaluation actuelle, notre but est d'apporter des outils de catégorisation permettant aux officiers d'évaluer plus finement l'information.

MOTS-CLÉS: Renseignement, Évaluation de l'information, Admiralty Scale, Descriptif vs Évaluatif, Vague, Taxonomie.

*Cite as: B. Icard, Facts versus Interpretations in Intelligence: A Descriptive Taxonomy for Information Evaluation, in: O. Gapenne, O. Chopin (Eds.), *Intellectica*, Cognition and Intelligence, 78 (1), 89-105, 2023.

I - INTRODUCTION

One of the activities of intelligence agencies consists in evaluating the informational messages obtained from human sources [e.g. [NATO AJP-2.1, 2016](#)] through the collection of “*human intelligence*”, also simply called “HUMINT” for HUMAN INTelligence [[NATO AJP-2.3, 2021](#)]. At the step of “Information Evaluation”, an intelligence officer uses a 6×6 scale known as the Admiralty System, or NATO System, to grade the credibility of message contents and the reliability of their sources, based on six levels of evaluation in both cases. The evaluation of credibility and reliability is made independently but then, gradings are crossed to give an overall evaluation of the message [[NATO STANAG-2511, 2003](#), [DIA-2, 2010](#), [NATO AJP-2.1, 2016](#)].

In addition to this scoring procedure, officers are also required to always distinguish objective facts from their subjective interpretations of them [see [NATO STANAG-2511, 2003](#), p. 2]. Despite this requirement, various experiments have shown that officers usually do not follow such a recommendation that we may call the *facts versus interpretations recommendation* [e.g. [Baker et al., 1968](#), [Kelly and Peterson, 1971](#), [Johnson, 1973](#)]. The most common explanation is that they are confused about the inner meaning of the scale dimensions (e.g. opacity of the labels, conflicts between those labels and their corresponding descriptions), their distinct levels of evaluation, and the resultant scores. Not surprisingly, these confusions also materialize empirically by a lack of agreement within and between intelligence raters [see e.g. [Wark, 1964](#), [Baker et al., 1968](#), [Johnson, 1973](#), [Samet, 1975](#)].

This paper proposes to make the intelligence recommendation operative by shedding light on the more objective, i.e. descriptive, side of the scale. To this end, we track the descriptive parameters which control the evaluative dimensions of the scale. Consistent with the STANAG doctrine, we postulate that officers rely on descriptive dimensions of *truth* and *honesty* to rate *credibility* and *reliability* and strike resultant scores. In line with doctrinal instructions, we see credibility as a subjective opportunity to estimate the objective truth of message contents based on the evidence available in the context of the elements to be evaluated. By contrast, reliability is seen as a way to evaluate the honesty of the source based on its informational pedigree.

Extracting the dimensions of truth and honesty underlying the scale is certainly essential. But further distinctions can be made to enforce the *facts versus interpretations recommendation*. Empirical findings show that regarding the rating labels, linguistic directionality can be used to simplify the 6×6 evaluative scale (see section 1 for a description) into a 3×3 matrix [[Budescu et al., 2003](#), [Teigen and Brun, 1995, 1999](#), [Mandel et al., 2022](#)]. In particular, [Mandel et al. \[2022\]](#) have observed that officers use the positive versus negative directionality of the credibility and reliability labels to group them into three distinct categories. Credibility and reliability ratings are separated into positive versus negative groupings, while an extra grouping corresponds to cases in which the evaluation of credibility and reliability is neither positive nor negative but uncertain.

In this paper, we move from groupings at the evaluative level to groupings at the descriptive level. We obtain a 3×3 descriptive matrix that distinguishes three levels of truth for contents (“*True*”, “*False*”, “*Indeterminate*”) and three levels of honesty for sources (“*Honest*”, “*Dishonest*”, “*Imprecise*”). The two levels correspond to clear-cut descriptive dimensions: “*True*” and “*False*” concerning contents, and “*Honest*” and “*Dishonest*” for sources. Extra dimensions correspond to areas of uncertainty concerning the truth of the content (“*Indeterminate*”) and the honesty of the source (“*Imprecise*”). Those areas of uncertainty will be analyzed extensively through the lenses of linguistic vagueness, using in particular Égré & Icard’s subdivisions between *semantic indeterminacy* and *pragmatic imprecision* (see [Égré and Icard, 2018](#), based on [Lasersohn, 1999](#), [Kennedy, 2007](#), [Solt, 2015](#)).

In section II, we first present the alphanumeric scale, or matrix, defined by the Admiralty Code for information evaluation (subsection II-1). We then put more emphasis on the *facts versus interpretations recommendation* that officers have to follow when they use the scale to evaluate information

(subsection II-2). We also put more emphasis on the strengths of this extant scale for the evaluation task (subsection II-3).

Our purpose in section III is to show that intelligence officers fail to separate objective facts from their subjective interpretations when they use the scale. We start by pointing out some inner limitations in the definition of the scale dimensions, ratings and scores (subsection III-1). Based on past empirical results, we argue that these limitations tend to generate inconsistencies within and between officers which prevent them from distinguishing facts from interpretations (subsection III-2). Finally, we make a proposal to help enforce the *facts versus interpretations recommendation*. This proposal consists in extracting the descriptive scale, more precisely the objective dimensions of truth and honesty, that remain hidden in the background of the Admiralty Code (subsection III-3).

In section IV, we set out our descriptivist proposal in detail. We first group the credibility and reliability ratings by directionality to obtain a 3×3 evaluative scale (subsection IV-1). Then, we extract the descriptive version of this simplified scale, by considering the dimensions of truth for contents and of honesty for sources (subsection IV-2). Finally, we introduce a 3×3 taxonomy of intelligence message which integrates clear-cut categories (*true* and *false* contents, *honest* and *dishonest* sources) and borderline cases based on *semantic vagueness* — when the content is semantically indeterminate, or corresponding to *pragmatic vagueness* — when the source is less informative than expected. Finally, we link each descriptive category to a more familiar label (*information*, *misinformation*, *lie*, etc.) (subsection IV-3).

We conclude in section V by summarising the main steps of our proposal. We point out the practical benefits of using a taxonomy of intelligence messages in addition to the original alphanumeric scale of the Admiralty Code. But we also draw attention to potential refinements of this taxonomy and to further integrations.

II - INFORMATION EVALUATION IN INTELLIGENCE

1. The traditional scale for information evaluation

Message		Content Credibility					
		1	2	3	4	5	6
Source Reliability	A	A1	A2	A3	A4	A5	A6
	B	B1	B2	B3	B4	B5	B6
	C	C1	C2	C3	C4	C5	C6
	D	D1	D2	D3	D4	D5	D6
	E	E1	E2	E3	E4	E5	E6
	F	F1	F2	F3	F4	F5	F6

Table 1: The 6×6 Traditional Alphanumeric Matrix.

In the Intelligence Cycle [TTA, 2001, DIA-2, 2010],¹ one stage regarding HUMINT consists in exploiting the raw information collected by human sources to derive meaningful intelligence.²

¹Different versions of the intelligence cycle exist (see TTA, 2001, DIA-2, 2010 for the specificities of the four-stages French model; see ODN, 2011 for the five-stages American model). However, all these versions agree on the existence of a specific step devoted to information evaluation.

²Naturally, this stage of Processing is not restricted to HUMINT. It also extends to the other forms of collected intelligence such as open-source intelligence (OSINT), but also more “technical” forms, such as signals intelligence (SIGINT), imagery intelligence (IMINT), measurement and signatures intelligence (MASINT), — with specificities in all those cases.

In this stage usually called “Processing” (or “Exploitation”), the step of “Information Evaluation” is of crucial importance. A competent authority, more precisely an intelligence officer, rates “*an item of information in respect of the reliability of the source, and the credibility of the information*” [e.g. [NATO STANAG-2511, 2003](#), [FM-2-22.3, 2003](#), [DIA-2, 2010](#), [NATO AJP-2.1, 2016](#)]. To perform this task, the officer uses a 6×6 alphanumeric scale historically known as the Admiralty System (or Admiralty Code) since its development in World War II (WWII) by the Naval Intelligence Division of the British Admiralty [see [McLachlan, 1968](#), [Montagu, 1977](#)]. For a given informational message, the officer is asked to rate dimensions of credibility and reliability separately [e.g. [NATO STANAG-2511, 2003](#), p. 2], and to mark a resultant score based on those two ratings.

<i>Content Credibility</i>	<i>Source Reliability</i>
1: Confirmed	A: Completely Reliable
2: Probably True	B: Usually Reliable
3: Possibly True	C: Fairly Reliable
4: Doubtful	D: Not Usually Reliable
5: Improbable	E: Unreliable
6: Cannot Be Judged	F: Cannot Be Judged

Table 2: Linguistic Labels for the Ratings.

Tables 1 and 2 present the alphanumeric matrix according to the NATO STANAG-2511: the six linguistic labels associated with both credibility and reliability ratings, and the thirty-six resultant scores.³ The conventional reading of score “E1”, for example, is that the source of the message is judged as “Unreliable” while the content of the message is “Confirmed” by other information. It should be noted that, strictly speaking, credibility and reliability ratings range from 1 to 5 and from A to E (respectively). Ratings 6 and F are not evaluatives *stricto sensu* since in those cases, external evidence is lacking to cross-check contents and sources.

Credibility ratings are captured through different labels that express *decreasing levels of confidence in the truth* of the message given external evidence consistent (or inconsistent) with its content being true (*contextual evidence*). Degree 1 corresponds to an absolute label (“Confirmed”) that captures cross-checked certainty. Degrees from 2 to 5 correspond to adverbial modulations (“Probably True”, “Possibly True”, etc.) which capture *high consistency* (“Probably True”), *moderate consistency* (“Possibly True”), *weak inconsistency* (“Doubtful”) and *blatant inconsistency* (“Improbable”). Degree 6 is ascribed when *no evidential ground* exists for assessing the credibility of the information.

Reliability ratings are captured through adverbial modulations of the evaluative term “*reliable*” (“Completely Reliable”, “Usually Reliable”, etc.). They correspond to *decreasing levels of confidence in the honesty* of the source based on its informational pedigree (*historical evidence*). Reliability is conceived here as a two-dimensional concept that aggregates the source’s tendency to be *truthful* (i.e. to regularly provide objectively true information), but also to be *honest* (i.e. to regularly provide information it believes to be true). But in this paper, analyses of reliability only rely on the source’s honesty since its truthfulness is controlled by the truth or falsity of the contents it delivers. The exact definitions of each linguistic label, with minor variations, are provided in the doctrine manuals [see e.g. [NATO STANAG-2511, 2003](#)].

³In the other doctrine reports [e.g. [FM-2-22.3, 2003](#), [DIA-2, 2010](#), [NATO AJP-2.1, 2016](#)], minor differences with the STANAG-2511 concern the linguistic labels of the ratings (for example, “Doubtfully True” is used instead of “Doubtful”), but the credibility and reliability dimensions, and the positive versus negative directionality of their labels, are identical. We leave these specificities aside: they have no impact on our own proposal.

2. The facts versus interpretations recommendation

The 6×6 alphanumeric matrix is followed by a recommendation we may label “*facts versus interpretations*” and that is expressed in the NATO Standardization Agreement as follows [see [NATO STANAG-2511, 2003](#), p. 2]:

“Intelligence reports transmit facts and/or assessments. The distinction between fact and interpretation must always be clearly indicated.”

Here the reference to “*intelligence reports*” indicates that the recommendation extends to other steps within the processing stage, in particular to the step of analysis.⁴ But in the specific case of information evaluation, the recommendation acts as a guideline for using the 6×6 matrix properly. The recommendation is methodological : officers are expected to distinguish facts from interpretations and to report this distinction carefully. From a conceptual or, more precisely, epistemological perspective, this distinction coincides with the separation between objective and subjective perspectives on information. Facts correspond to objective reality : they are independent from the officer’s perception and cognitive insights. By contrast, interpretations are subjective : they refer to the officer’s personal reading of the world based on its expertise and cognitive abilities.

So, in the context of information evaluation, the *facts versus interpretations recommendation* is rooted in the distinction between objective versus subjective perspectives on information. But since the scale aims to provide evaluations of information, the subjective perspective is more accurately evaluative in this case. By contrast, the opposite perspective is more precisely descriptive than objective. This descriptive perspective requires categorizing messages in an objective sense, independently, or at least separately, of one’s subjective evaluation. Accordingly, we choose to read the *facts versus interpretations recommendation* as a distinction between the descriptive versus evaluative perspectives on information.

Nonetheless, we may wonder whether the alphanumeric scale allows such a distinction in practice. As a matter of fact, various empirical results show that this is not the case. Before reviewing those results, we highlight some positive features of the extant alphanumeric scale.

3. Some virtues of the alphanumeric scale

Firstly, the alphanumeric scale is *balanced*. In fact, dimensions of credibility and reliability cover both semantic and pragmatic features of informational messages. The semantic level of messages is taken into account by credibility ratings which aim to evaluate the truth or falsity of message contents. The pragmatic aspect of messages concerns the sources’ communicative intentions and are evaluated through reliability ratings. This semantic/pragmatic parity conforms with Paul Grice’s intention-based semantics according to which both semantic and pragmatic aspects contribute to the meaning of messages [[Grice, 1975](#)].

The existing scale also relies on *relevant* dimensions. Credibility and reliability provide useful information for assessing the quality of a given message. Credibility is a crucial parameter for believing a content on a secure basis while reliability is an essential criterion for trusting a source or not. Sometimes called “*veracity*” or “*accuracy*”, credibility refers to the *plausibility* that a message is true with respect to a given context of evidence. Reliability refers to the officer’s relative *certainty* that a given source is honest with respect to its informational pedigree (i.e. its disposition to tell what it believes to be the truth over time).

Finally, the scale is *sensitive* since officers are given five levels for evaluating the credibility of the content (“Confirmed”, “Probably True”, etc.) and five levels for evaluating the reliability of the source (“Reliable”, “Fairly Reliable”, etc.), as well as one extra level on each scale when no

⁴It is also expected that intelligence analysts who make assumptions and/or predictions to clearly separate facts from their more personal opinions [[DCDC JDP 2-00, 2011](#), p. 3-28; [MCTP 2-10B, 2019](#), p. 4-4].

evaluation can be made (“Cannot Be Judged”). The intelligence scale being a 6×6 matrix, officers can choose between 36 combinations to evaluate the quality of intelligence messages.

Despite those positive aspects, the scale has also many limitations. Based on past empirical findings, we now argue that the rating labels, their respective levels, and the resultant scores are confusing and lead to inconsistencies within and between raters. As a consequence, officers cannot comply with the *facts versus interpretations recommendation* by using the current scale.

III - DISTINGUISHING FACTS FROM INTERPRETATIONS

1. Intrinsic limitations in the features of the scale

Semantic confusion can be observed at three levels in the extant scale: *credibility ratings*, *reliability ratings* and their combination into *scores*. First, labels and descriptions associated with *credibility ratings* are conflicting. Labels are based on *plain* adverbial quantifiers (“Probably True”, “Possibly True”, etc.) while the corresponding descriptions define them as *conditional* adverbial quantifiers. According to descriptions, the existence of consistent or inconsistent evidence is conditional on the ascription of those plain ratings. In addition to that, a problematic concept of corroboration underlies those various descriptions. For instance, as [Capet and Revault d’Allonnes \[2013\]](#) point out, the highest level of credibility, namely **1**, is labeled “Confirmed” as if independent confirmation by other source(s) had established absolute certainty in the truth of the message. But without knowing that the source currently evaluated is (also) “Completely Reliable”, no absolute certainty can be ascribed to the message itself [see [DIS, 2001](#), [Capet and Revault d’Allonnes, 2013](#), p. 115].

Second, the inner meaning of *reliability ratings* is also confusing. Their labels and intended descriptions are ambiguous. It can refer to: (a) the trust the officer puts into the source of the message; (b) the credibility that the source itself attributes to the message; and (c) the proper understanding by the source of the message it is delivering. In addition to this ambiguity, descriptions associated with reliability labels are opaque. They are based on complex notions of *authenticity* and *trustworthiness* without providing minimal definitions of them or explaining how they combine to form reliability ratings.

Third, semantic confusion is inherent to *resultant scores* since they attempt to cross credibility and reliability ratings whose descriptions are themselves confused. In addition to that, it seems that resultant scores are deprived of proper meanings and semantic references: they do not *fuse* but simply *cross* credibility and reliability ratings to give an overall insight into sources and contents. As a consequence, resultant scores do not have proper *extensions* and *intensions* on their own, and their semantic reference is reduced to that of the initial credibility and reliability ratings.

This lack of proper meaning generates confusion and inconsistency regarding the linguistic directionality of some resultant scores. In psycholinguistics, directionality is the tendency of some linguistic terms, in particular adjectives or adverbs, to point toward a *positive* or a *negative* direction [e.g. [Moxey and Sanford, 1986](#), [Teigen and Brun, 1995, 1999](#), [Budescu et al., 2003](#)]. For instance, adjectives “*probable*”, “*possible*” and the adverb “*likely*” point toward a *positive* direction while “*impossible*”, “*doubtful*” or “*unlikely*” point toward a *negative* direction. But though the directionality of all the ratings is clear and univocal, — either positive (e.g. “Probably” in credibility rating **2** or “Usually” in reliability rating **B**) or negative (e.g. “Doubtful” in rating **4** or “Unreliable” in rating **E**), the directionality of resultant scores is sometimes confused since it may combine ratings with opposite directionality. For instance, the directionality of score **A5** is problematic since rating **A** (“Completely Reliable”) is positively directional while rating **5** (“Improbable”) is negatively directional. We observe similar kinds of inconsistency with other scores crossing grades with opposite directionality (e.g. **E1**, **D1**, **E2**, **A4**, etc.).

2. Subjective inconsistencies over the scale interpretation

2.1. Confusion within intelligence officers

Rather expectedly, this confusion gives rise to inconsistencies within officers. Experimental results show that officers vary in their interpretations of the ratings across contexts. Many authors observed that the alphanumeric scale is not used at full range by intelligence officers, thus indicating that they have difficulties understanding the ratings and resultant scores [see e.g. [Phelps et al., 1980](#), for a critical review]. In particular, [Baker et al. \[1968\]](#) analyzed 695 rating reports made by American army officers during field exercises. Only 40% of those reports contained ratings of both credibility and reliability. And when both evaluations were made, 87% of the scores ($N = 608$ out of 695) fell strictly along the diagonal of the scale, that is on the continuum **A1-B2-C3-D4-E5-F6**. Later replicated by [Samet \[1975\]](#), those results show that intelligence raters, despite not considering credibility and reliability as identical dimensions,⁵ do not make a clear-cut difference between the two since they see them as strongly correlated and overlapping. Because of this redundancy, officers do not take full advantage of the evaluative opportunities of the scale. Not only did the large majority of ratings fall on the diagonal, but score **B2** alone⁶ actually received 75% of all the total number of ratings (i.e. $N = 518$ out of 695).⁷

On a more abstract level, confidence judgments also reveal asymmetries in the officers' respective understandings of the ratings and scores. When raters had to express their relative confidence in the resultant scores they obtained, [Meeland and Rhyne \[1967\]](#) observed that the 36 possible scores were not equally weighted by intelligence personnel. In fact, officers turned out to be six times more confident of a **B1** rating than of a **F3** rating. Some explanations have been proposed to account for this discrepancy. One of them is that the higher the credibility rating of a message content, the more corroborating evidence officers have for cross-checking it and the more confident they are that the message is true. By contrast, the lower the credibility rating of a message content, the less corroborating evidence officers have for verifying it and the less confident they are that the message is true. Similar arguments have been provided for explaining the discrepancy between confidence judgments and reliability ratings [see [Peterson, 2008](#), on this point].

2.2. Inconsistencies between intelligence officers

In addition to within-officers confusion, between-officers inconsistencies have also been noticed. From a qualitative perspective, adverbial quantifiers (“*probably*”, “*possibly*”, “*usually*”, “*fairly*”, etc.) turn out to be interpreted differently by intelligence personnel. [Baker et al. \[1968\]](#) noticed high inconsistencies in the responses elicited by different officers in similar contexts. Looking at the ratings provided during an intelligence course, they found that those ratings differed from the instructor's solution about 49% of the time concerning credibility, and 15% of the time concerning reliability.

Following those observations, researchers have proposed to match adverbial quantifiers with quantitative encodings such as percentages, probabilities, odds, etc. Concerning credibility ratings in particular, a theoretical hypothesis is that officers would assign the probability ranges and absolute means given in Table 3.

⁵See [Samet \[1975\]](#) and also [Miron et al. \[1978\]](#) on the relative importance of both dimensions, as well as others, during information evaluation. See [Icard \[2019\]](#) on the logical implementation of previous experimental results showing that credibility is the main dimension of the evaluation while reliability only plays an adjustment role.

⁶viz. “Usually Reliable”-“Probably True”.

⁷Some proposals were made to help officers use the scale at full range, in particular by [Baker and Mace \[1973\]](#) who proposed to assist officers with a decision flow chart (viz. a sequence of basic clear-cut questions), but no strong improvement was observed with this approach.

Credibility Ratings	Probability Ranges	Probability Means
1	0.80 - 1.0	0.90
2	0.60 - 0.80	0.70
3	0.40 - 0.60	0.50
4	0.20 - 0.40	0.30
5	0 - 0.20	0.10

Table 3: Expected Probability Degrees for Credibility Ratings.

We can compare those theoretical encodings with the numerical probabilities that the first Director of the CIA’s Office of National Estimates, Sherman Kent, has proposed for intelligence quantifiers (see Kent, 1964). He made this proposal for the alphanumeric scale in particular, and for credibility ratings 2 (“Probably True”), 3 (“Possibly True”) and 5 (“Improbable”) more specifically. Except for rating 3 (“Possibly True”), the theoretical probabilities given in Table 3 are consistent with the ones Kent proposed. For example, he matched the adjective “probable” with a probability range of 0.63 - 0.87 (absolute mean: 0.75) and the adjective “*improbable*” with a probability range of 0.20 - 0.40 (absolute mean: 0.30). By contrast, he matched the adjective “possible” with a wider range of probability, that is to more than 0 but less than 1.0, which corresponds to a range of 0 - 1.0 and to an absolute probability mean of 0.50. But are these degrees and ranges empirically settled?

Collected results only concern the assignment of absolute probability degrees. They are consistent with expected probabilities but give rise to high variability between officers [see Levine and Eldredge, 1970, Kelly and Peterson, 1971, Samet, 1975, Teigen and Brun, 1999]. Although Wark [1964] observed that the modal adverb “*probably*” corresponds to a probability degree of 0.75 with a very high consensus (90% of inter-agreement), he observed only 53% of inter-agreement that chances are about even (0.50) in case of the modal adverb “*possibly*”. This lack of consensus was also observed by Johnson [1973] for the adjective “*possible*” whose mean probability was 0.62 but with results varying from 0.04 to 0.80 across officers. The consensus observed by Wark in case of “*probably*” was not replicated by Johnson for adjective “probable” whose results varied from 0.10 to 0.99 across officers (mean: 0.51). Johnson also tested the adjective “*improbable*”, which was assigned an absolute degree of 0.17 but with results varying from 0 to 0.70.

Such variability in the officers’ probabilistic interpretations of adverbial expressions shows the vagueness induced by qualitative vocabulary.⁸ Outside intelligence studies, other fields reveal similar inter-individual inconsistencies. In linguistics, for instance, Lichtenstein and Newman [1967] collected results that were consistent with the ones of Wark: they observed consistency for the adjective “*probable*” (as well as for the adverb “*usually*”), and variability for “*possible*” [see also Budescu and Wallsten, 1985]. In the field of medicine, a comparison of studies from O’Brien [1989] and Bryant and Norman [1980] also show consistency for adjective “probable” that turned out to be associated with a probability of 0.75 in O’Brien’s study and with a probability of 0.77 in Bryant & Norman’s study [see also Hobby et al., 2000]. But their studies disagree over adjective “possible” that was associated with a probability degree of 0.25 in O’Brien’s study and to a degree of 0.47 in Bryant & Norman’s. Besides that, O’Brien and Bryant & Norman obtained similar rates for “*probable*” to those of Wark for “probably” ($p = 0.75$), but results were different for “*possible*” in case of O’Brien ($p = 0.25$).

In the field of intelligence, proposals for such numerical encodings remain a major challenge due to the officers’ varying interpretations of them. But for the sake of argument, suppose that different officers would agree on a single interpretation for each probability degree (or range) and thus, that probabilities could be completely substituted to intelligence adverbs. Quite unfortunately, other issues would materialize that concern people’s asymmetric perception of risk and probabilities. For instance, Pighin et al. [2011] observed that for two distinct numerical probabilities $1/307 < 1/28$ where $1/307$ is the probability for a child to have Down syndrome while $1/28$ is

⁸In their critical review of within-officers inconsistencies, Phelps, Halpin, Johnson & Moses talk of the “*loose, ambiguous language used to communicate uncertain intelligence information*” [Phelps et al., 1980, p. 1].

the probability for a child to have insomnia, Down syndrome was interpreted as more likely to happen than insomnia on a 7-point scale ranging from “extremely low” to “extremely high”. This effect which is qualified elsewhere as a “Severity Bias” [e.g. Weber and Hilton, 1990, Bonnefon and Villejoubert, 2006], reveals that the interpretation of probabilities is not purely *extensional* but strongly linked to *expected utilities*, namely to the *severity of the outcomes* involved. Égré [2014] and then Égré and Cova [2015] made similar observations concerning *numbers*. Based on experiments about the vague determiner “many”, they observed that cardinals, like probabilities, are not interpreted from a sole *extensional* perspective (based on a given comparative class) but are also judged *intensionally* based on moral expectations of desirability.⁹

We have seen that officers’ inconsistencies first rely on the pragmatic interpretation of the scale whose dimensions are semantically confused (subsubsection III-2.1). But we have also seen that substituting numerical encodings to intelligence adverbs will create new inconsistencies in the officers’ interpretations of the numerals (subsubsection III-2.2). For this reason, we will continue using qualitative adverbs as evaluative dimensions for intelligence ratings. We now argue that the inconsistencies observed within and between officers make the *facts versus interpretations recommendation* currently impossible to apply.

2.3. Failure of meeting the facts versus interpretations recommendation

For officers in charge of information evaluation, the precondition for respecting the *facts versus interpretations recommendation* is to be able to mark the distinction along the extant scale. But to do so, officers should be able to point towards both sides of the distinction when they use the scale. However, reaching this goal implies that the scale allows a clear and consistent understanding of what *facts* are, on the one hand, and of what *interpretations* are, on the other hand. We argue that this is not the case.

First, expressing one’s subjective interpretations is already problematic with the current scale. The scale evaluative dimensions are intended to help officers give subjective interpretations on messages. But the lack of *intra-* and *inter-*agreement over these dimensions shows that consistent interpretations cannot be settled with the extant scale. Second, identifying objective facts through the lenses of the scale is doomed to failure. Currently, the scale only makes *explicit* the evaluative dimensions used for its assessment role: credibility for the content of the message versus reliability for its source. But the more factual, that is to say descriptive, dimensions underpinning those evaluative dimensions remain strictly *implicit* in the extant scale: *truth* of the message content in case of credibility, and *honesty* of the message source in case of reliability. Since those descriptive dimensions are hidden in the background of the scale, officers cannot use it to delineate objective facts from their subjective interpretations.

We now describe our own proposal to make more room for the descriptive perspective during information evaluation. Without replacing the extant scale, this proposal only seeks to help officers respect the *facts versus interpretations recommendation*.

3. A new proposal to identify intelligence facts

First, following previous works on directionality [Mandel et al., 2022, more specifically], we argue that linguistic labels used for separating levels of credibility and of reliability have some directionality, positive or negative, that can be used to simplify the 6×6 alphanumeric scale into a 3×3 scale. The scale obtained after simplification allows to distinguish intelligence messages whose

⁹Égré & Cova tested people’s ascriptions of the vague quantifier “many” in the context of a fire involving children. They asked participants to agree whether “many” could be ascribed in case exactly 5 children (out of 10) escaped the fire versus in case exactly 5 children (out of 10) perished in the fire. In this case, people were more prone to judging that “many children perished in the fire” than that “many children escaped the fire”. Their findings parallel those of Knobe showing that moral considerations have an increasing effect on people’s ascriptions of intentions [Knobe, 2003].

evaluation is clear-cut (clearly positive or clearly negative), from intermediate messages whose evaluation is more borderline (neither clearly positive nor clearly negative).

The second part of our proposal follows the intuition that respecting the facts versus interpretations recommendation implies to be able to extract the descriptive matrix behind the evaluative scale. So far, the 3×3 matrix relies on dimensions that are interpretative and non-factual, i.e. credibility and reliability. We propose to extract the descriptive dimensions of truth and honesty that remain hidden in the background of those two dimensions. As a result, we obtain a 3×3 descriptive matrix based on three levels of truth for contents and three levels of honesty for sources. As before, two levels are clear-cut while an extra level concerns borderline cases, — the one for truth, the other for honesty. Based on Égré & Icard’s taxonomy of vagueness [see [Égré and Icard, 2018](#)], the distinction between semantic and pragmatic vagueness is used to see more clearly into those borderline cases.

In the third part of our proposal, we use the nine categories naturally derived from this 3×3 descriptive matrix to elaborate a taxonomy of intelligence messages. We show that those categories can be matched to more familiar linguistic labels: *information*, *misinformation*, *subjective lie*, *objective lie*, *omission*, *dissimulation*, etc. This way, we seek to provide officers with well-founded and meaningful labels since they are based on well-known expressions in the literature. But we also aim to overcome potential inconsistencies observed in the directionality of resultant scores which are based on crossing grades with opposite directionality (e.g. **A5** for instance).

It is important to note that our 3×3 descriptive scale is not meant to replace the evaluative scale. Though our scale is extracted from the 6×6 alphanumeric matrix, the former is descriptive while the latter is evaluative. Our descriptive matrix is only intended to help officers better separate facts from interpretations in intelligence. Truth and honesty being factual instead of evaluative dimensions, intelligence officers are expected to be more consistent when using the descriptive scale than they were with the evaluative scale.

IV - A TAXONOMY OF INTELLIGENCE MESSAGES

1. The directionality of the evaluative ratings

Let us go back to the evaluative dimensions of the 6×6 scale given in Table 2 (see subsection II-1). Different experimental results have shown that numerical adjectives and adverbial quantifiers, are perceived as having either a positive or a negative directionality [[Teigen and Brun, 1995, 1999](#), [Budescu et al., 2003](#), [Irwin and Mandel, 2019](#)]. For instance, Teigen & Brun ran a study with a total of 62 students (32 American, 30 Norwegian) to determine whether 24 verbal probability expressions directed the students’ foci of attention toward a positive or a negative evaluation. Amongst those expressions, some were identical, or very close to, the verbal expressions used in the intelligence ratings: e.g. “*absolutely certain*”, “*probable*”, “*possible*”, “*doubtful*”, “*improbable*”. In line with intuitions, it turned out that 100% of the subjects that were tested for expressions “*absolutely certain*”, “*probable*” and “*possible*” rated those expressions as positively directional while 100% of the subjects that were tested for “*doubtful*” and “*improbable*” rated them as negatively directional [see [Teigen and Brun, 1995](#), study 1 for details].

Recently, [Mandel et al. \[2022\]](#) devoted two experiments to the phenomenon of directionality in the context of information evaluation. In particular, they aimed to better understand how directionality influences the interactions between numerical probabilities associated with resultant scores (called *information accuracy* according to their terminology) and numerical probabilities associated with credibility and reliability ratings. Consistent with Teigen & Brun’s preliminary results [[Teigen and Brun, 1995](#)], they observed that credibility ratings from 1 to 5 and reliability ratings from A to E can be separated into positive versus negative groupings. Graphically, indeed, a clear inflection appears between the mean probabilities associated with credibility ratings 1, 2 and 3

whose directionality is judged as clearly *positive*,¹⁰ and the mean probabilities associated with ratings 4 and 5 whose directionality is seen as clearly *negative*.¹¹ Similarly, they observed a clear inflection between the mean probabilities of reliability ratings: A, B and C whose directionality is perceived as clearly *positive*¹² form a separate group from ratings D and E whose directionality is seen as clearly *negative*.¹³ In our proposal, we adopt the same clear-cut distinction between positive versus negative groupings of the intelligence ratings.

Both for credibility and reliability, however, we may wonder about the role of intermediary areas located between those two groupings. The theoretical existence of such neutral areas, and their practical usefulness, seems natural for expressing one's uncertainty in the context of information evaluation. But the extant scale does not provide any specific ratings for officers to report when they are *uncertain* whether credibility should be rated positively (as in ratings 1 to 3) or negatively (as in ratings 4 and 5), or for when they are *uncertain* whether reliability should be rated positively (as in ratings from A to C) or negatively (as in E and F). As a matter of fact, data collected by Mandel et al. [2022] tend to show that those intermediary areas are associated with rating "Cannot Be Judged" both in case of credibility (rating 6) and in case of reliability (rating F). Looking more closely at the inflection points of their data, the mean probabilities associated with ratings 6 and F fall exactly between the probabilities of positively directional ratings, on the one hand, and the probabilities of negatively directional ratings, on the other hand. This indicates that professionals seem to use ratings 6 and F, i.e. "Cannot Be Judged", to evaluate neutral areas in case of credibility and reliability. In the present article, we do not discuss this identification between neutral areas and ratings "Cannot Be Judged", but we also assume the existence of such neutral groupings in case of credibility and reliability.

Let us summarize our progress before going any further. Based on the groupings observed by intelligence researchers, we can now simplify the 6×6 scale into a 3×3 (still) evaluative matrix. More precisely, we can isolate 3 distinct levels for credibility ratings (positive, negative and neutral) and 3 distinct levels for reliability ratings (positive, negative and neutral). That being done, we can now transform this simplified 3×3 evaluative scale into a 3×3 descriptive matrix. To this end, let us look closer at the connections between the ratings and underlying dimensions of truth and honesty they evaluate.

2. Truth and honesty as descriptive dimensions of the scale

We start out by focusing on the two clear-cut groupings of the 3×3 evaluative scale. In the first grouping, the officer is confident that the credibility of the message or the reliability of its source is *positive* while in the second grouping, the officer is confident that credibility or reliability is *negative*. By looking at those groupings from a *descriptive point of view*, this means that the officer judges the content to be *true* or the source to be *honest* in the first grouping, while it judges the content to be *false* or the source to be *dishonest* in the second grouping.

Let us now put our attention on the neutral areas of the 3×3 evaluative scale. In the neutral area for credibility, the officer is uncertain whether the content has a positive or negative credibility while in the neutral area for reliability, the officer is uncertain whether the source has a positive or negative reliability. From an evaluative perspective, both areas express the officer's epistemic uncertainty about credibility and reliability. But *from a descriptive perspective*, those areas express the officer's uncertainty about the *truth status of the message content* in case of credibility, and about the *communicative intents of the source* in case of reliability.

As a matter of fact, uncertainty concerning contents and sources can be seen as manifestations of a same phenomenon called *linguistic vagueness* [see Égré, 2018]. In the broad sense, vagueness occurs when the meaning of an expression, or word, is indeterminate due to a one-many relationship

¹⁰i.e. "Confirmed", "Probably True", "Possibly True".

¹¹i.e. "Doubtful", "Improbable".

¹²i.e. "Completely Reliable", "Usually Reliable", "Fairly Reliable".

¹³i.e. "Not Usually Reliable", "Unreliable".

between the expression and its set of potential interpretations [Russell, 1923]. The interpretation of a vague expression is then compatible with an open-ended range of possible meanings [Pinkal, 1995, Raffman, 2013]. But amongst vague expressions, a distinction can be made between vagueness which occurs at the level of message contents, i.e. with respect to their semantic meaning, and vagueness which occurs at the level of sources, i.e. with respect to their pragmatic intents. According to the literature, the first manifestation of vagueness concerns the truth status of message contents and is called *semantic vagueness*, or *indeterminacy*; while the second manifestation concerns the level of cooperativity of sources and is called *pragmatic vagueness*, or *imprecision* [see Lasersohn, 1999, Kennedy, 2007, Solt, 2015].

More precisely, semantically vague expressions have intrinsically uncertain truth conditions, either because they have one free dimension of interpretation, as in gradable adjectives like “tall” or “long” concerned by *degree-vagueness* (terminology from Alston, 1964, see also Burks, 1946, Kennedy, 2007), or because those expressions have many free dimensions of interpretation, as in multidimensional adjectives like “rich”, “intelligent” or “big” now concerned by *combinatorial vagueness* (terminology from Alston, 1964, see also Waismann, 1945, Burks, 1946). By contrast, pragmatically vague expressions have definite truth conditions but they can be used with slack depending of the context. Pragmatic vagueness can take two forms: *generality vagueness* [Fine] when it relies on using underspecific expressions such as “some”, “most”, “always”; or *approximation* [Lasersohn, 1999] when it consists in modifying precise expressions (e.g. “around”, “about”, “nearly”, “roughly”) or in using an expression with a coarser meaning, like in the sentence “France is hexagonal” [see Austin, 1962, Lewis, 1970].

Following all those distinctions, Égré and Icard [2018] have proposed a taxonomy in which expressions can be either precise (e.g. numbers, locations, proper names) or vague. In case they are vague, expressions are either *semantically vague* or *pragmatically vague*. Then, semantic vagueness separates into *degree-vagueness* (one free dimension of interpretation) and *combinatorial vagueness* (many free dimensions of interpretation); while pragmatic vagueness is divided into *generality* (underspecificity) and *approximation* (modifying precise expression, coarsening).

It is important to note that since then, Égré & Icard’s taxonomy has been refined [Guélorget et al., 2021, Icard et al., 2022]. On top of extant delineations, linguistic expressions are also classified as being either *factual* or *subjective* depending on whether there are used to report objective information or more personal opinions [Kennedy, 2013]. Factual information can be communicated through precise expressions (numbers, locations, proper nouns, etc.) but also through pragmatic imprecision. Indeed, pragmatically vague expressions are considered *factual* since they have semantically precise truth conditions, although they can be interpreted with slack. By contrast, *subjective* information is associated with semantically vague expressions since those expressions have uncertain truth conditions and can be interpreted in multiple ways [Solt, 2018, Kaiser and Wang, 2021].

In the present paper, however, we do not use this factual versus subjective distinction since we are mainly interested in the officer’s perspective on information. By contrast, the factual versus subjective distinction is made from the speaker’s perspective: it concerns the way speakers, in particular intelligence sources, use linguistic vagueness to communicate factual versus subjective information to addressees, in particular to intelligence officers. But here, we are mainly focused on the way officers categorize messages *objectively* or interpret them more *subjectively*. Though the source and officer’s perspectives interact, our objective/subjective distinction, and its refinement into the descriptive/evaluative distinction, only concerns officers in this article.

Let us now illustrate cases of semantic and pragmatic vagueness in the context of information evaluation. In case of credibility, we know that the officer’s uncertainty is based on the fact that the content of the message is borderline between truth and falsity. This is a case of *semantic vagueness* since the truth status of the message is indeterminate: it can be true in some respect, that is with respect to some interpretation, but false with respect to some other interpretation. By way of

illustration, we consider the dialogue in (1) between an intelligence officer and its source. The source who belongs to a belligerent country has some revelations to make about a new submarine program in its country.

- (1) **Officer:** “What are the strengths of the submarines?”
Source: “They are powerful.”

In (1), the source’s answer can be understood in multiple ways and thus, it is difficult for the officer to interpret it in one way or another. In this case, semantic vagueness relies on the *combinatorial vagueness* of the adjective “powerful” [Sassoon, 2013]: submarines may be powerful *in some respect* (because they are able to launch nuclear missiles for example) but not powerful *in all respects* (because their speed is limited compared to other extant vessels). That is: “powerful” is multidimensional and can be sub-interpreted when understood as “*in some respect*”, or super-interpreted when understood as “*in all respects*” [Alxatib and Pelletier, 2011]. As a consequence, the source’s answer fails to be unambiguously true, contrary to the requirement of the Gricean Supermaxim of Quality: “*Try to make your contribution one that is true*” [Grice, 1975]. In the next section, we will see that such semantic vagueness can be used as a double-edged sword [Égré and Icard, 2018], either to inform the officer, when the source is honest, or to deceive it, when the source is dishonest.

After analyzing vagueness with respect to credibility, let us consider reliability. Now, the officer’s uncertainty is due to the fact that the source is borderline between being honest and being dishonest. This is *pragmatic vagueness*, or *imprecision*, since the officer cannot evaluate the accuracy of the source’s statement based on the context and its informational pedigree. To better understand vagueness of this kind, we consider the dialogue in (2) between the officer and its source on the submarine program.

- (2) **Officer:** “Are the ten submarines already operational?”
Source: “Most of them.”

In (2), the source provides an answer which is less informative than it is required by the Gricean first Maxim of Quantity: “*Make your contribution as informative as is required (for the current purposes of the exchange)*” [Grice, 1975]. In this case, pragmatic vagueness relies on the *generality vagueness* of the adverb “most”. Contrary to “powerful” in (1), vagueness in (2) does not mean any indeterminacy since “most” can receive determinate truth conditions relative to a fixed countable domain. But by answering “most”, the source fails to be maximally informative by being underspecific in its response to the officer. Specifying an exact number of submarines that are *operational* would be more informative in that respect.

3. Providing meaningful labels to message types

Once extracted the descriptive dimensions of the scale, we obtain a 3×3 descriptive matrix based on two dimensions: “*Truth of the Content*” vs “*Honesty of the Source*”. We distinguish three levels of truth (“*True*”, “*False*”, “*Indeterminate*”) and three levels of honesty (“*Honest*”, “*Dishonest*”, “*Imprecise*”). Concerning truth, a content is said to be *true* if it corresponds to objective facts, *false* if it does not and *indeterminate* if the content is semantically vague. Concerning honesty, sources are *honest* in case they tell exactly what they believe to be true, *dishonest* if they tell the opposite of what they believe and *imprecise* if they are pragmatically vague. Dimensions of truth and honesty combine in a 3×3 descriptive matrix whose general framework is presented in Table 4. In the next subsections, we describe the different categories, or message types *t*, of this matrix and associate meaningful labels to them.

Message type <i>t</i>		Truth of the Content		
		<i>True</i>	<i>Indeterminate</i>	<i>False</i>
Honesty of the Source	<i>Honest</i>	t_1	t_5	t_2
	<i>Imprecise</i>	t_7	t_9	t_8
	<i>Dishonest</i>	t_3	t_6	t_4

Table 4: The 3×3 Descriptive Matrix behind the 6×6 Evaluative Scale.

3.1. Classical types of messages

Classically, the content of a message is either *true* or *false* while its source is either *honest* or *dishonest*. No intermediate option is available. According to Table 5 that focuses only on the classical part of Table 4, four categories of messages can be identified in the classical sense: $t_1 = (\text{Honest/True})$, $t_2 = (\text{Honest/False})$, $t_3 = (\text{Dishonest/True})$ and $t_4 = (\text{Dishonest/False})$.

Message type <i>t</i>		Truth of the Content		
		<i>True</i>	<i>Indeterminate</i>	<i>False</i>
Honesty of the Source	<i>Honest</i>	t_1 = <i>information</i>		t_2 = <i>misinformation</i>
	<i>Imprecise</i>			
	<i>Dishonest</i>	t_3 = <i>subjective lie</i>		t_4 = <i>objective lie</i>

Table 5: The Classical Message Types of the 3×3 Descriptive Matrix.

We first consider message types t_1 and t_2 in which the source is clearly *honest*. In both types, the source (subjectively) believes the content of its message to be true. But in type t_1 , the content is objectively *true* while in type t_2 , the content is objectively *false*. So contrary to type t_1 , the source itself is mistaken by the content it delivers in type t_2 : it believes it as true but the content is false in reality. With respect to Grice’s norms of communication [Grice, 1975], the source is cooperative in both t_1 and t_2 . The source also respects the four Gricean Maxims, except that it violates the Supermaxim of Quality (“Try to make your contribution one that is true”) in case of type t_2 . Following

common labels, we choose to call type t_1 “information”, and type t_2 “misinformation”.

In classical epistemology, information is defined as “well-formed, meaningful and veridical data” [see Floridi, 2007, p. 31]. Since *veridical* means *truthful* here, information is conceived as true data conveyed by truthful sources. Consistent with this view, Dretske states that “false information, misinformation (...) are not varieties of information” [Dretske, 1983, p. 57] while for Grice: “false information is not an inferior kind of information; it just is not information” [Grice, 1989, p. 371]. Since Fallis [2009] and Floridi [2011], however, false information is considered as information *per se*. They call *misinformation* “well-formed and meaningful data (i.e. semantic content) that are false” [Floridi, 2011, p. 260]. Unlike their predecessors, they consider information to be logically independent from truth or falsity. Although the default position is that information is true, as it is the case in type t_1 , information can also be false. For this reason, we chose to call type t_2 “misinformation” with the aim to avoid confusion with t_1 , traditionally called “information”.

By contrast, we now address types t_3 and t_4 in which the source is clearly *dishonest*. In both cases, the source is non-cooperative and even deceitful: it delivers information it clearly *believes to be false*. As a result, the source inevitably violates Grice’s first Maxim of Quality in both t_3 and t_4 . But in type t_3 , the information the source delivers is only subjectively believed to be false: it is in fact objectively *true*. In type t_4 , by contrast, not only is the information subjectively believed to be false, it is also objectively *false*. So the source breaches the Gricean Supermaxim of Quality (“Try to make your contribution one that is true”) in type t_4 but not in type t_3 . For that reason, type t_3 happens to be less misleading than type t_4 . Even though the source is dishonest in both cases, it provides epistemically worse information in the second case than in the first.

According to the traditional definition of lying, a lie is a believed-to-be-false statement that a speaker makes to an addressee with the intention that the latter believes its statement to be true [e.g. Augustine, 395, Williams, 2010, Mahon, 2015, Wiegmann et al., 2016]. In this so-called “subjective account of lying”, it is sufficient that the speaker believes its statement to be false for it to count as a lie. Contrary to a more recent view known as the “objective account” [Krishna, 1961, Grotius, 2005, Turri and Turri, 2015, Turri, 2021], the subjective definition does not take objective falsity as a necessary condition for lying. To mark this distinction in our own taxonomy, we call type t_3 “subjective lie” since it refers to a dishonest statement that is true, while we call type t_4 “objective lie” since it corresponds to a dishonest statement that is false.

In addition to those classical types based on binary contents (either *true* or *false*) and binary sources (either *honest* or *dishonest*), we now extend our investigations to types based on intermediate dimensions with respect to contents (“Indeterminate”) or to sources (“Imprecise”). In both cases, vague expressions are a double-edge sword: they can be used either to inform or to deceive the officer.

3.2. Some types based on semantic vagueness

Table 6 focuses on vagueness regarding informational contents, as in $t_5 = (\text{Honest/Indeterminate})$ and $t_6 = (\text{Dishonest/Indeterminate})$. Those types are instances of *semantic vagueness*, or *indeterminacy*: in both cases, the content of the message is borderline between truth and falsity since it may be true in some respects but false in others. The difference is that the source of the message is *honest* in type t_5 and *dishonest* in type t_6 .

With respect to the Gricean norms, the source violates the Supermaxim of Quality (“Try to make your contribution one that is true”) in both t_5 and t_6 since its message fails to be *clearly*, or *unambiguously*, true. Following Égré & Icard’s subdivision of semantic indeterminacy between one-dimensional and multi-dimensional vagueness, the source also violates Grice’s Second Maxim of Manner (“Avoid ambiguity — i.e., avoid language that can be interpreted in multiple ways”) in types t_5 and t_6 , in particular when the content of those messages involve multidimensional vague adjectives. The main difference between t_5 and t_6 is that the source respects the first Maxim of Quality

Message type t		Truth of the Content		
		True	Indeterminate	False
Honesty of the Source	Honest		t_5 = <i>error-avoidance</i>	
	Imprecise			
	Dishonest		t_6 = <i>half-truth</i>	

Table 6: Types based on Semantic Vagueness in the 3×3 Scale.

(“Do not say what you believe to be false”) in t_5 but not in t_6 . In the first case, the source is *honest* and does not provide information it believes to be false (no matter whether what the source tells is objectively true or false). By contrast, the source violates the first Maxim of Quality in t_6 since the source is *dishonest* and delivers information it explicitly believes to be false.

To better understand the distinction between types t_5 and t_6 , let us go back to example (1). In this case, the officer asks information about the strengths of the new submarines and the source answers that they are “powerful”. Here the officer cannot be sure of the sense of “powerful” the source has in mind. In fact, “powerful” can be used either to inform or to deceive the officer. In the first case which would correspond to type t_5 , vagueness is used honestly: the source says “powerful” instead of a more precise expression to prevent the officer from believing something false, for instance that the submarines carry nuclear bombs when this is not the case. But in a completely different situation now corresponding to type t_6 , the source can perfectly use vagueness dishonestly in order to deceive the officer. In this case, the source says “powerful” to make the officer infer a false conclusion, for example that the engines of the submarines are very fast when this is false.

We decide to label type t_5 “error-avoidance” to express cases in which the source is honest and relies on vagueness to prevent potential deception. By contrast, following Égré & Icard’s terminology, we call type t_6 “half-truths” to express situations in which the source is dishonest and expects the officer to be deceived.

3.3. Some other types based on pragmatic vagueness

By contrast with Table 6, Table 7 focuses on vagueness regarding the sources’ intents, as in $t_7 = (\text{Imprecise/True})$ and $t_8 = (\text{Imprecise/False})$. Those types are instances of *pragmatic vagueness*, or *imprecision*: in both cases, the source is borderline between honesty and dishonesty since it is neither clearly honest nor clearly dishonest. But the difference between those two types is that the content of the message is *true* in type t_7 and *false* in type t_8 .

Let us compare t_7 and t_8 based on the Gricean norms. First of all, the source breaches the first Maxim of Quantity (“Make your contribution as informative as is required (for the current purposes of the exchange)”) in both types since the source is pragmatically imprecise and fails to be maximally informative to the officer. However, the source does not breach the first Maxim of Quality (“Do not say what you believe to be false”) in any of those types. Although the source does not say what

Message type t		Truth of the Content		
		True	Indeterminate	False
Honesty of the Source	Honest			
	Imprecise	t_7 = <i>omission</i>		t_8 = <i>dissimulation</i>
	Dishonest			

Table 7: Types based on Pragmatic Vagueness in the 3×3 Scale.

it believes in t_7 and t_8 (i.e. that the message is true in t_7 and false in t_8), the source actually does not say anything it believes to be false (in this case: that the content is false in t_7 and true in t_8). Finally, setting aside what the source believes, it violates the Supermaxim of Quality (“*Try to make your contribution one that is true*”) in both t_7 and t_8 since the source being underspecific, it withholds information about the exact state of the world in both types. But in t_7 , the source *conceals the truth* of the message content while in t_8 , the source is *underspecific with respect to the falsity* of the message.

To see more clearly into the distinctions between t_7 and t_8 , let us go back to example (2). In (2), the officer asks the source whether the 10 expected submarines are operational and the source answers that “*most*” of them are already fully operational. To distinguish t_7 from t_8 along this example, let us consider two situations. In the first situation, suppose that *exactly* 8 out of the 10 submarines are operational. In this case which would correspond to type t_7 , the source’s answer with “*most*” is true. But its answer is underspecific: the source relies on the pragmatic vagueness of “*most*” to *hide the truth* about the *exact number* of already operational submarines (i.e. 8 out of 10). In the second situation, suppose that only 5 out of the 10 submarines are operational. In this case now corresponding to type t_8 , the source’s answer with “*most*” is false, but this answer is not as blatantly false as it would have been the case by answering with an exact false number, for instance “*9 out of 10*”. By contrast with type t_7 based on hiding the truth, type t_8 is based on using vagueness to hide something more explicitly false. In the situation we describe, answering “*9 out of 10*” would be more explicitly false than just answering “*most*” since the former expression is more precise than the latter is.

From a more epistemological perspective, types t_7 and t_8 correspond to different *epistemic goals* [Chisholm and Feehan, 1977, Fallis, 2014, 2018]. In case of type t_7 , the source aims to prevent the officer from acquiring a true belief when it hides the truth by using vague language. In type t_8 , by contrast, the source aims to prevent the officer from correcting a false belief when it uses vagueness to hide that something is false. To mark this distinction, we label t_7 as “*omission*” (of the truth) and t_8 as “*dissimulation*” (of the false).

Finally, t_9 is an unusual type mixing semantic indeterminacy *with* pragmatic imprecision: $t_9 =$ (*Imprecise/Indeterminate*). In case of t_9 , the content of the message has unclear truth conditions while its source is less informative than required by the Gricean first Maxim of Quantity. We will not describe this type in detail since t_9 is a simple compound of the more primitive types t_5 , t_6 ,

t_7 and t_8 . We choose to label type t_9 as “mixed” to insist on this specificity. Table 8 summarizes the whole descriptive taxonomy of intelligence messages we have presented.

Message type t		Truth of the Content		
		True	Indeterminate	False
Honesty of the Source	Honest	t_1 = <i>information</i>	t_5 = <i>error-avoidance</i>	t_2 = <i>misinformation</i>
	Imprecise	t_7 = <i>omission</i>	t_9 = <i>mixed</i>	t_8 = <i>dissimulation</i>
	Dishonest	t_3 = <i>subjective lie</i>	t_6 = <i>half-truth</i>	t_4 = <i>objective lie</i>

Table 8: A Descriptive Taxonomy of Intelligence Messages.

V - CONCLUSION

Many experiments have shown that intelligence raters cannot respect the *facts versus interpretations recommendation* when they use the 6×6 Admiralty scale. To make the recommendation effective, we have proposed to isolate the dimensions of truth and honesty on which the evaluative dimensions of the scale rely. Based on those descriptive dimensions, we have introduced a 3×3 taxonomy of intelligence messages, the aim of which being to help officers separate facts from interpretations and to make more acute evaluations of information.

We have insisted on the fact that this descriptive proposal is not meant to replace the evaluative procedure. We simply believe that having a descriptive taxonomy of messages brings some advantages to the traditional approach. First, since truth and honesty are descriptive dimensions, they are less prone to multiple interpretations than evaluative dimensions, and this makes our descriptive scale easier to grasp and manipulate than the evaluative scale. Second, the descriptive scale offers three levels of discrimination instead of six. Though more delineations are possible, in particular concerning intermediate dimensions, having fewer levels may help officers take more resolute decisions when they have to categorize messages. Third, the labels associated with descriptive categories (e.g. *misinformation*, *subjective lie*, *objective lie*, etc.) provide more information than the initial degrees of truth and honesty they result from. Those labels are not redundant with the initial degrees and they offer meaningful categories for officers to use. Finally, the descriptive and evaluative scales complement one another. Since the evaluative ratings aim to appreciate dimensions of truth and honesty on a subjective basis, the alphanumeric scale can be seen as a tentative procedure to ascribe the descriptive categories introduced in the taxonomy.

We consider two directions for future work. The first direction concerns the granularity of our taxonomy as well as its potential extension. Refining the taxonomy could help discriminate between subcases of semantic vagueness (*degree-vagueness*, *combinatorial vagueness*), and between subvarieties of pragmatic imprecision (*approximation*, *generality*). In addition to that, further work with Égré and other collaborators has led to link pragmatic and semantic vagueness with the communication of factual versus subjective information [see e.g. Guélorget et al., 2021, Icard et al., 2022].

But this link, which concerns the speaker’s perspective on information, has not been followed in the present paper, which focuses on the addressee’s perspective, the one of the officer. That being said, both perspectives interact in information evaluation, as they do in any communicative exchange, so an adequate model of evaluation should be more integrative in that respect.

The second research direction we are considering is to test the efficacy of the 3×3 taxonomy to help enforce the *facts versus interpretations recommendation*.¹⁴ We want to test whether this taxonomy is a valid descriptive subscale of the 6×6 evaluative scale. But we also want to know whether our taxonomy is sufficiently reliable to be used by officers with confidence. That being done, a follow-up study could consist in looking more closely into intermediate areas within both scales. We should clarify the relationship between areas seen as *neutral* by officers in the 6×6 scale, and dimensions of the 3×3 scale related to vague information. Recent results show that ratings 6 and F (“Cannot Be Judged”) may provide evaluations of such neutral areas [see Mandel et al., 2022], even though those ratings are classically assumed as non-evaluative [see e.g. NATO STANAG-2511, 2003]. But the interaction between ratings “Cannot Be Judged” and the evaluation of vagueness has yet to be determined.

ACKNOWLEDGMENTS

Funding support for this work was provided by the RAPID project DIEKB (DGA01D19018444), the ANR program HYBRINFOX (ANR-21-ASIA-0003), and the ANR-17-EURE-0017 (FrontCog). We thank Denis Bonnay, Hans van Ditmarsch, Percy Kemp, Marie-Jeanne Lesot, Raphaël Millière, Jakob Süskind and Louis Vayssette for very useful comments on an earlier version of this article, and also two anonymous referees for suggestions and corrections on the present version. We are particularly grateful to Paul Égré for having stimulated and inspired many ideas and connections made in the present paper, and also to Philippe Capet for our continuous exchanges on the topics of deception and intelligence processing.

REFERENCES

- William P. Alston. *Philosophy of Language*. Prentice Hall, 1964.
- S. Alxatib and J. Pelletier. On the psychology of truth-gaps. In R. Nouwen, R. van Rooij, U. Sauerland, and H-C. Schmitz, editors, *Vagueness in communication*, pages 13–36. Springer, LNAI 6517, 2011.
- Saint Augustine. On lying. *St. Augustine: Treatises on Various Subjects*. New York: *The Fathers of the Church*, pages 47–109, 395.
- John L. Austin. *Sense and Sensibilia*. Oxford University Press, 1962.
- James D. Baker and Douglas J. Mace. Certitude judgments revisited. *Unpublished manuscript, US Army Research Institute for the Behavior and Social Sciences*, 1973.
- James D. Baker, James M. McKendry, and Douglas J. Mace. Certitude judgments in an operational environment. *US Army Technical Research Note*, 200, 1968.
- Jean-François Bonnefon and Gaëlle Villejoubert. Tactful or doubtful? expectations of politeness explain the severity bias in the interpretation of probability phrases. *Psychological Science*, 17(9): 747–751, 2006.
- Geoffrey D. Bryant and Geoffrey R Norman. Expressions of probability: words and numbers. *The New England Journal of Medicine*, 302(7):411, 1980.

¹⁴See Mandel [2020] on the importance of testing intelligence methodologies.

- David V Budescu and Thomas S Wallsten. Consistency in interpretation of probabilistic phrases. *Organizational Behavior and Human Decision Processes*, 36(3):391–405, 1985.
- David V Budescu, Tzur M Karelitz, and Thomas S Wallsten. Predicting the directionality of probability words from their membership functions. *Journal of Behavioral Decision Making*, 16(3): 159–180, 2003.
- Arthur W. Burks. Empiricism and vagueness. *The Journal of Philosophy*, 43(18):477–486, 1946.
- Philippe Capet and Adrien Revault d’Allonnes. La cotation dans le domaine militaire: doctrines, pratiques et insuffisances. In Philippe Capet and Thomas Delavallade, editors, *L’évaluation de l’information: confiance et défiance*. Hermès, 2013.
- Roderick M. Chisholm and Thomas D. Feehan. The intent to deceive. *Journal of Philosophy*, 74(3): 143–159, 1977.
- DCDC JDP 2-00. *Understanding and Intelligence Support to Joint Operations*. Shrivenham, UK: Development, Concepts and Doctrine Centre, 2011.
- DIA-2. Doctrine interarmées. *CICDE, Renseignement d’intérêt militaire et contre-ingérence (RIM & CI)*, 2, 2010.
- DIS. Intelligence wing student précis. *Defence Intelligence, Security Center*, 2001.
- Fred I. Dretske. Précis of knowledge and the flow of information. *Behavioral and Brain Sciences*, 6 (1):55–90, 1983.
- Paul Égré. Intentional action and the semantics of gradable expressions: (on the knobe effect). In Bridget Copley and Fabienne Martin, editors, *Causation in Grammatical Structures*, pages 176–205. Oxford University Press, 2014.
- Paul Égré. *Qu’est-ce que le vague?* Librairie philosophique J. Vrin, 2018.
- Paul Égré and Florian Cova. Moral asymmetries and the semantics of many. *Semantics and Pragmatics*, 8:13–1, 2015.
- Paul Égré and Benjamin Icard. Lying and vagueness. In Jörg Meibauer, editor, *Oxford Handbook of Lying*. Oxford University Press, 2018.
- Don Fallis. A conceptual analysis of disinformation. *iConference 2009 Papers*, 2009.
- Don Fallis. The varieties of disinformation. In *The Philosophy of Information Quality*, pages 135–161. Cham: Springer, 2014.
- Don Fallis. Lying and omissions. In Jörg Meibauer, editor, *Oxford Handbook of Lying*. Oxford University Press, 2018.
- Kit Fine. Vagueness, truth and logic. *Synthese*, 30.
- Luciano Floridi. In defence of the veridical nature of semantic information. *European Journal of Analytic Philosophy*, 3(1), 2007.
- Luciano Floridi. *The Philosophy of Information*. Oxford University Press, 2011.
- FM-2-22.3. *Human Intelligence: collector operations*. Department of the United States Army, 2003.
- Herbert P Grice. Logic and conversation. In *Speech acts*, pages 41–58. Brill, 1975.
- Paul H. Grice. *Studies in the way of words*. Cambridge: Harvard University Press, 1989.
- Hugo Grotius. *The rights of war and peace*, 3 vols. Indianapolis: Liberty Fund, 2005.

- Paul Guélorget, Benjamin Icard, Guillaume Gadek, Souhir Gahbiche, Sylvain Gatepaille, Ghislain Ateazing, and Paul Égré. Combining vagueness detection with deep learning to identify fake news. In *Proceedings of 24th International Conference on Information Fusion*, page 8, 2021.
- Jonathan L. Hobby, BD Tom, C Todd, PW Bearcroft, and Adrian K Dixon. Communication of doubt and certainty in radiological reports. *The British Journal of Radiology*, 73(873):999–1001, 2000.
- Benjamin Icard. *Lying, deception and strategic omission : definition and evaluation*. Theses, Université Paris sciences et lettres, February 2019. URL <https://theses.hal.science/te1-02170022>.
- Benjamin Icard, Ghislain Ateazing, and Paul Égré. Vago: un outil en ligne de mesure du vague et de la subjectivité. In *Conférence Nationale sur les Applications Pratiques de l'Intelligence Artificielle (PFIA 2022)*, pages 68–71, 2022.
- Daniel Irwin and David R Mandel. Improving information evaluation for intelligence production. *Intelligence and National Security*, 34(4):503–525, 2019.
- Edgar M. Johnson. Numerical encoding of qualitative expressions of uncertainty. *ARI Technical Paper*, 250, 1973.
- Elsi Kaiser and Catherine Wang. Packaging information as fact versus opinion: Consequences of the (information-) structural position of subjective adjectives. *Discourse Processes*, pages 1–25, 2021.
- C. W. Kelly and C.R. Peterson. Probability estimates and probabilistic procedures in current intelligence analysis. *Report, International Business Machines Corporation*, 71-5047, 1971.
- Christopher Kennedy. Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and philosophy*, 30(1):1–45, 2007.
- Christopher Kennedy. Two sources of subjectivity: Qualitative assessment and dimensional uncertainty. *Inquiry*, 56(2-3):258–277, 2013.
- Joshua Knobe. Intentional action and side effects in ordinary language. *Analysis*, 63(3):190–194, 2003.
- Daya Krishna. 'lying' and the compleat robot. *The British Journal for the Philosophy of Science*, 12(46):146–149, 1961.
- Peter Lasersohn. Pragmatic halos. *Language*, 75(3):522–551, 1999.
- Jerrold M. Levine and Donald Eldredge. The effects of ancillary information upon photointerpreter performance. *Report, American Institutes for Research*, 70-14, 1970.
- David Lewis. General semantics. *Synthese*, 22(1/2):18–67, 1970. ISSN 00397857, 15730964. URL <http://www.jstor.org/stable/20114749>.
- Sarah Lichtenstein and Robert J. Newman. Empirical scaling of common verbal phrases associated with numerical probabilities. *Psychonomic Science*, 9(10):563–564, 1967.
- James E Mahon. The definition of lying and deception. *The Stanford Encyclopedia of Philosophy*, 2015.
- David R. Mandel. The occasional maverick of analytic tradecraft. *Intelligence and National Security*, 35(3):438–443, 2020. doi: 10.1080/02684527.2020.1723830. URL <https://doi.org/10.1080/02684527.2020.1723830>.
- David R. Mandel, Daniel Irwin, Mandeep Dhama, and David Budescu. Meta-informational cue inconsistency and judgment of information accuracy: Spotlight on intelligence analysis. *Journal of Behavioral Decision Making*, 11 2022. doi: 10.1002/bdm.2307.

- Donald McLachlan. *Room 39: Naval intelligence in action 1939-45*. Weidenfeld & Nicolson, 1968.
- MCTP 2-10B. *MAGTF Intelligence Production and Analysis*. Washington, D.C.: Headquarters United States Marines Corps, 2019.
- T. Meeland and R. F. Rhyne. A confidence scale for intelligence reports: an application of magnitude estimation scaling. *Technical Note, Stanford Research Institute*, 4923-31, 1967.
- Murray S. Miron, Samuel Patten, and Stanley M. Halpin. The structure of combat intelligence ratings. *US Army Technical Paper*, 286, 1978.
- Ewen Montagu. *Beyond Top Secret U*. Peter Davies Ltd, 1977.
- Linda M Moxey and Anthony J Sanford. Quantifiers and focus. *Journal of semantics*, 5(3):189–206, 1986.
- NATO AJP-2.1. Allied joint doctrine for intelligence procedures. *North Atlantic Treaty Organization*, edition B, version 1:1–92, 2016.
- NATO AJP-2.3. Allied joint doctrine for human intelligence (humint). *North Atlantic Treaty Organization*, 2021.
- NATO STANAG-2511. Standardization agreement, intelligence report. *North Atlantic Treaty Organization*, 2511:1–15, 2003.
- Bernie J O'Brien. Words or numbers? the evaluation of probability expressions in general practice. *JR Coll Gen Pract*, 39(320):98–100, 1989.
- ODN. National intelligence: A consumer's guide. *Office of the Director of National Intelligence*, 2011.
- Joshua J Peterson. Appropriate factors to consider when assessing analytic confidence in intelligence analysis. *Master of Science Thesis, Department of Intelligence Studies, Mercyhurst College, Erie, Pennsylvania*, 2008.
- Ruth H. Phelps, Stanley M. Halpin, Edgar M. Johnson, and Franklin L. Moses. Implementation of subjective probability estimates in army intelligence procedures: A critical review of research findings. *Research Report US Army*, 1242, 1980.
- Stefania Pighin, Jean-François Bonnefon, and Lucia Savadori. Overcoming number numbness in prenatal risk communication. *Prenatal diagnosis*, 31(8):809–813, 2011.
- Manfred Pinkal. *Logic and Lexicon: a Study of the Indefinite*. Dordrecht: Springer, 1995.
- Diana Raffman. *Unruly words: A study of vague language*. Oxford University Press, 2013.
- Bertrand Russell. Vagueness. *The Australasian Journal of Psychology and Philosophy*, 1(2):84–92, 1923.
- Michael G. Samet. Quantitative interpretation of two qualitative scales used to rate military intelligence. *Human Factors*, 17(2):192–202, 1975.
- Galit W Sassoon. A typology of multidimensional adjectives. *Journal of semantics*, 30(3):335–380, 2013.
- Stephanie Solt. Vagueness and imprecision: Empirical foundations. *Annu. Rev. Linguist.*, 1(1): 107–127, 2015.
- Stephanie Solt. Multidimensionality, subjectivity and scales: Experimental evidence. *The semantics of gradability, vagueness, and scale structure: Experimental perspectives*, pages 59–91, 2018.

- Karl Halvor Teigen and Wibecke Brun. Yes, but it is uncertain: Direction and communicative intention of verbal probabilistic terms. *Acta Psychologica*, 88(3):233–258, 1995. ISSN 0001-6918. doi: [https://doi.org/10.1016/0001-6918\(93\)E0071-9](https://doi.org/10.1016/0001-6918(93)E0071-9). URL <https://www.sciencedirect.com/science/article/pii/0001691893E00719>.
- Karl Halvor Teigen and Wibecke Brun. The directionality of verbal probability expressions: Effects on decisions, predictions, and probabilistic reasoning. *Organizational behavior and human decision processes*, 80(2):155–190, 1999.
- TTA. *Traité toutes armes. Renseignement*, (150), 2001.
- Angelo Turri and John Turri. The truth about lying. *Cognition*, 138:161–168, 2015.
- John Turri. Objective falsity is essential to lying: an argument from convergent evidence. *Philosophical Studies*, 178(6):2101–2109, 2021.
- Friedrich Waismann. Open texture. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 19:119–150, 1945.
- David L Wark. The definition of some estimative expressions. *Studies in Intelligence*, 8(4):67–80, 1964.
- Elke U. Weber and Denis J Hilton. Contextual effects in the interpretations of probability words: Perceived base rate and severity of events. *Journal of Experimental Psychology: Human Perception and Performance*, 16(4):781, 1990.
- Alex Wiegmann, Jana Samland, and Michael R Waldmann. Lying despite telling the truth. *Cognition*, 150:37–42, 2016.
- Bernard Williams. Truth and truthfulness. In *Truth and Truthfulness*. Princeton University Press, 2010.