Information Acquisition for Decision Making: Information Source Models

David Grace
Lehigh University

Eugene Perevalov
Lehigh University

Report: 12T-003

Information Acquisition for Decision Making: Information Source Models

Draft

Abstract

The optimal decision making problem in situations characterized by uncertainty and availability of information sources is considered in a general setting. This gives rise to the need for a quantitative framework for the description of information exchange between the decision maker and information sources. Two companion papers established the first two parts of such a framework: the concepts of question difficulty and answer depth. This paper explores the third component of the framework, i.e. quantitative models of information sources. The concept of a source model is introduced and several different models are proposed. The source model parameters and the pseudo-temperature function on the problem parameter space characterizing question difficulty and answer depth in the overall "ideal gas" information exchange model can be estimated from the observed source performance on a set of sample questions. Optimization based methods for such estimation are discussed.

1 Introduction

The problem of optimizing the process of additional information acquisition for general decision making problems has not received sufficient attention in research literature in spite of an existing need for such a methodology in a number of practical decision making situations. The latter are typically characterized by a high degree of uncertainty in the initial input data and an existence of a number of potential information sources that are believed to possess different "pieces" of additional information which are nevertheless difficult to make use of simply because the information they possess is not sufficiently well organized and "concentrated" in order to be readily used in, for instance, a stochastic optimization formulation. One could say that the information typical sources

possess is present largely in a latent form and needs to be activated, or actualized, to become useful in a mathematical formulation.

The need for a quantitative methodology for optimizing the additional information acquisition process for decision making problems of the sort described above motivates development of an analytical theory of an interaction between decision making problems and information sources. Given a decision making problem and and information source, the main task is to find a way to use the information source optimally from the viewpoint of maximum increase in the quality of the solution of the problem. A natural way to achieve that is to ask the information source question(s) about the input data for the problem. Clearly, the effect of the source's answer to a given question will depend on both the specific nature of the question (that can be more or less relevant to the problem) and the ability of the source to answer this question well (that can depend on how hard this particular question is for the source). So the decision maker's task is to find a question that would be simultaneously relevant to the given problem and be easy enough for the given source that the latter can answer it with high accuracy. This task can be described as that of optimal "alignment" between the problem and the information source. It is clear that completion of this task necessitates development of a quantitative framework describing question, answers and information sources as a logical first step.

The theory of questions and answers, respectively, was developed in [25] and [24]. The corresponding quantitative measures – question difficulty and answer depth – were proposed and general expressions for them derived using an axiomatic approach. The present article's focus is on quantitative models of information sources. The main goal of such models is to make it possible to evaluate a source's ability to answer various questions so that the problem of optimal "alignment" between a source and the given problem could be addressed.

1.1 Related Work

The concepts of information theory, which grew out of Shannon's work on communication theory, arise in many disciplines and the list of successful applications include, but is certainly not limited to, a simple derivation of statistical physics laws [16, 17], new algorithms in computer vision [29], new methods of analysis in climatology [23, 28], physiology [20] and neurophysiology [5]. This paper intends to continue the effort of applying information theoretic concepts to new areas and

applications, in particular that of decision making and optimization. For example, the concept of pseudo-energy, introduced in companion papers [25] and [24], builds on that of entropy and provides the foundation for the subject of this article – quantitative models of various information sources.

Evaluating a source's ability to answer various questions is closely related to the evaluation of probability forecasts by scoring rules. A scoring rule measures the accuracy of a forecast by computing a score based on how the forecast compares to the actual realization of the uncertain event. An early application of this is the Brier, or quadratic, score that evaluates probabilistic weather forecasts [4]. Scoring rules also provide an incentive for the forecaster to provide truthful probabilities and share a connection to subjective probability theory (e.g. [15, 27]. See [30, 1, 14] for a more thorough discussion of scoring rules and literature reviews. More closely aligned to work in this paper has been the development of scoring rules that also take into consideration the decision problem at hand, in particular [19]. They start with a decision problem and find scoring rules to fit the problem in a way that aligns interests of the expert and the decision maker. In contrast, to these scoring rules that measure the forecast with a single aggregated scalar value, our work introduces a pseudo-temperature function that evaluates the source over the entire state space. In this way, when there are multiple sources of information, the proper source can be chosen based on which one can more accurately answer the specified questions.

While the proposed methodology makes no assumption on what form the information source may take (only that it can provide an answer to a question), it is likely that the role of information source will likely be played by human experts in practical applications. This is also a common form for studying information sources throughout the literature. The problem of optimal usage of information obtained from human experts has been addressed mostly in the form of updating the decision maker's beliefs given a probability assessment from multiple experts [12, 13, 6, 7] and, in particular, optimal combining of expert opinions, including experts with incoherent and missing outputs [26]. Closely related to the approach initiated in this paper are the investigations on using and combining information of experts that partition the event differently [2] and on rules of updating probabilities based on outcomes of partially similar events [3]. The latter investigations essentially consider experts that provide qualitatively different information. The dependence of the quality of experts' output on the particular partition was also studied in [11]. In this papers and the companion papers [25, 24], the emphasis is on *optimizing* the particular type of information (i.e. partition) for the given expert(s) and the given decision making problem.

In the current work, models of information sources are considered which allows one to optimize not only the quantity of the acquired information but also its *content*. This is similar to the area of statistical decision making, where additional information is acquired to improve the decision quality. One can mention applications to innovation adoption [22, 18], fashion decisions [10] and vaccine composition decisions for flu immunization [21]. Typically, the amount of information in these applications is measured simply by the number of relevant observations of certain random variable realizations. Some authors introduced models (for instance, the effective information model) to account for the actual amount of information contained in the received observations [9, 8]. The common feature of this line of work is the search for an optimal trade-off between the amount of additional information obtained and the degree of achieving the original goal.

1.2 Outline

The rest of the article is organized as follows. In the next section, the main motivation for the framework developed in this paper is described. In section 3, some necessary preliminaries are given. In section 4, the overall framework and its three main components – questions, answers and information sources – are briefly reviewed, with the emphasis on information sources. In section 5, the concept of an information source model is introduced and several simple models are proposed. Section 6 describes the process of estimating – assuming the overall *ideal gas* question difficulty model – the pseudo-temperature function defined on the parameter space. Section 7 presents some numerical examples and section 8 gives a brief summary of the results.

2 Motivation: Decision Making Under Uncertainty

The main motivation for the proposed framework was discussed in the companion paper [25]. We briefly recap it here, for convenience. Uncertainty present in a decision making problem can often be described as a certain parameter space Ω that contains all possible sets of input data for the problem. The problem itself can be formulated as an optimization with respect to a suitably chosen criterion. One widely used criterion is that of optimizing an expected value of some objective function $f(\omega, x)$ that depends on both the input data ω and the decision x. The expectations is taken with respect

to a probability measure P that describes the information available to the decision maker:

$$\min_{x \in X} \mathbb{E}_P f(\omega, x) = \int_\Omega f(\omega, x) P(d\omega). \tag{1}$$

A notion of *loss* can usually be defined. It measures the performance of a solution obtained in the presence of uncertainty with respect to that of a solution that would have been obtained had the decision maker possessed the full information. For the formulation (1), the logical form of the loss is as follows.

$$L(P) = \int_\Omega f(\omega, x^*_P) P(d\omega) - \int_\Omega f(\omega, x^*_\omega) P(d\omega),$$

where x^*_P is a solution of (1) and x^*_ω is a solution of $\min_{x \in X} f(\omega, x)$ for the given input data ω.

An information source is assumed to be capable of answering questions concerning random outcomes on Ω. If we denote the set of all possible questions that the decision maker can ask the information source by $\mathcal{Q}$ and its answer to a particular question $Q \in \mathcal{Q}$ – by $A(Q)$, then the value of loss upon receiving the answer $A(Q)$ to the question Q becomes

$$L(P, Q, A(Q)) = \sum_a \Pr(A(Q) = a) \left(\int_\Omega f(\omega, x^*_{P_a}) P_a(d\omega) - \int_\Omega f(\omega, x^*_\omega) P_a(d\omega) \right), \tag{2}$$

where P_a is the measure on Ω conditional on reception of a particular value a of the answer A.

The goal of making optimal use of the information source can be formulated as that of finding a question Q such that an answer to it would allow the decision maker to find a solution with the minimum possible loss:

$$\min_{Q \in \mathcal{Q}} L(P, Q, A(Q)). \tag{3}$$

The question Q that achieves the minimum in (3) can be thought of as the optimal "link", or "channel" between the information source and the decision making problem.

3 Preliminaries

As was already stated in the companion paper [25], the basic ingredients of the proposed approach are the parameter space Ω equipped with a sigma-algebra $\mathcal{F}$ and a probability measure P that describes the initial information available to the decision maker.

If $C \in \mathcal{F}$ is a (measurable) subset of Ω, the conditional measure P_C on Ω is defined as

$$P_C(D) = \frac{P(D \cap C)}{P(C)}, \tag{4}$$

for arbitrary $D \in \mathcal{F}$.

A partition $\mathbf{C} = \{C_1, \ldots, C_r\}$ of Ω is a collection of (measurable) subsets $C_j \in \mathcal{F}$ of Ω such that $C_j \cap C_l = \emptyset$ for $j \neq l$ and $\cup_{j=1}^{r} C_j = \Omega$. A partition $\tilde{\mathbf{C}}$ is a *refinement* of $\mathbf{C}$ if every set from $\tilde{\mathbf{C}}$ is a subset of some set from $\mathbf{C}$. In such a case, $\mathbf{C}$ is a *coarsening* of $\tilde{\mathbf{C}}$.

If $\mathbf{C}' = \{C'_1, \ldots, C'_r\}$ and $\mathbf{C}'' = \{C''_1, \ldots, C''_s\}$ are two partitions of Ω then the partition $\mathbf{C} = \mathbf{C}' \cap \mathbf{C}''$ is defined as the partition that consists of all sets of the form $C'_i \cap C''_j$: $\mathbf{C}' \cap \mathbf{C}'' = \{C'_1 \cap C''_1, C'_1 \cap C''_2, \ldots, C'_r \cap C''_s\}$ (see Fig. 1 for an illustration). Clearly, $\mathbf{C}' \cap \mathbf{C}''$ is a refinement of both $\mathbf{C}'$ and $\mathbf{C}''$.

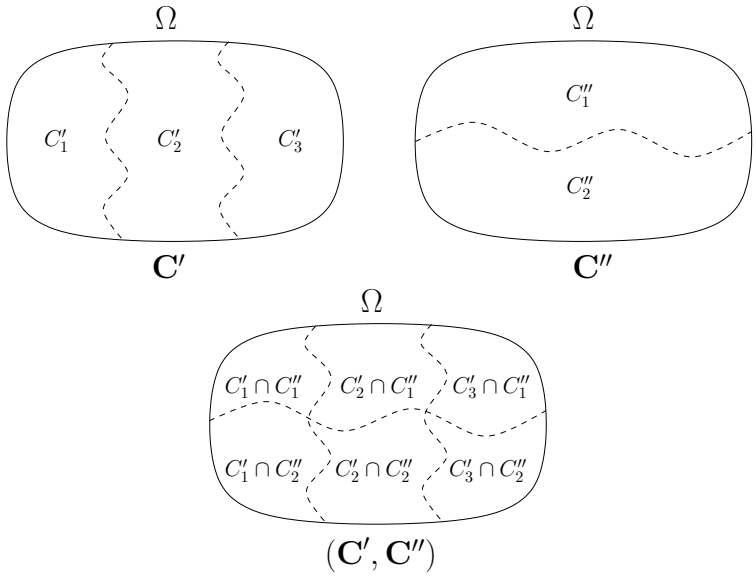

Figure 1: Two partitions of Ω and the corresponding joint partition.

If D is a subset of Ω and $\mathbf{C}' = \{C'_1, \ldots, C'_r\}$ is a partition of Ω, the partition $\mathbf{C}'_D = \{D \cap C'_1, \ldots, D \cap C'_r\}$ will be called the partition of D induced by the the partition $\mathbf{C}'$ of Ω (see Fig. 2).

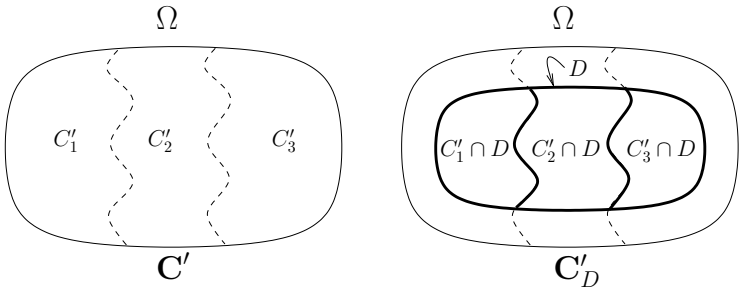

Figure 2: Partition $\mathbf{C}_D$ of set $D \subset \Omega$ induced by a partition $\mathbf{C}$ of Ω.

Besides standard partitions of Ω, we also make use of *incomplete* partitions $\mathbf{C} = \{C_1, \ldots, C_r\}$ such that $\cup_{i=1}^r C_i \neq \Omega$. For any partition $\mathbf{C}$, we use the notation $\hat{C} \equiv \cup_{i=1}^r C_i$. Clearly, partition $\mathbf{C}$ is complete if and only if $\hat{C} = \Omega$.

Given a complete partition $\mathbf{C} = \{C_1, \ldots, C_r\}$, the (initial) measure P can be decomposed into the corresponding conditional measures as follows.

$$P = \sum_{j=1}^{r} P(C_j) P_{C_j} \tag{5}$$

Let now $\mathbf{C}' = \{C'_1, \ldots, C'_r\}$ and $\mathbf{C}'' = \{C''_1, \ldots, C''_s\}$ be two incomplete partitions of Ω such that $\hat{C}' \cap \hat{C}'' = \emptyset$. The union partition $\mathbf{C} = \mathbf{C}' \cup \mathbf{C}''$ is defined as follows: $\mathbf{C} = \{C'_1, \ldots, C'_r, C''_1, \ldots, C''_s\}$, i.e. as a union of all subsets in the two constituent partitions. Clearly, $\mathbf{C}' \cup \mathbf{C}''$ is complete if and only if $\hat{C}' \cup \hat{C}'' = \Omega$. In case $\hat{C}' \cap \hat{C}'' \neq \emptyset$, the union partition $\mathbf{C}' \cup \mathbf{C}''$ is not defined.

4 Overall Framework: Main Components

The overall setup is discussed in the companion paper [25]. The main components involved are the information source, questions concerning random outcomes on Ω and its answers to these questions. While the articles [25] and [24] are mainly devoted to the study of questions and answers, respectively, the main focus of the present work is an investigation of the information source itself. To make the presentation self-contained, we include a brief discussion of the main results of references [25] and [24].

4.1 Questions

Given the parameter space Ω, a sigma-algebra $\mathcal{F}$ and an initial measure P on $(\Omega, \mathcal{F})$ (that we often refer to as just a measure on Ω), a question is identified with a partition $\mathbf{C} = \{C_1, C_2, \ldots, C_r\}$ of Ω that is allowed to be *incomplete*, i.e. the set $\hat{C} \equiv \cup_{j=1}^r C_j$ may be a proper subset of Ω. The questions for which $\hat{C} = \Omega$ are called *complete* or *multiple-choice* questions. The incomplete questions for which the corresponding partition consists of a single set $C \subset \Omega$ are called *free-response* questions, and incomplete questions with partitions consisting of several sets are called *mixed* questions. See [25] for additional details. Since a question is identified with the partition that describes it the terms "question" and "partition" are often used interchangeably.

A *difficulty function* $G(\Omega, \mathbf{C}, P)$ can be associated with any question $\mathbf{C}$. The particular form of $G(\Omega, \mathbf{C}, P)$ can be determined if some reasonable requirements, or, equivalently, *postulates*, are imposed. This was done in the companion paper [25] where a particular system of postulates that embodied *linearity* and *isotropy* properties of the difficulty function was proposed. The main theorem (stated here for complete questions) proved in [25] derives the general form of the difficulty function that is required to satisfy such postulates.

Theorem 1 *Let the function $G(\Omega, \mathbf{C}, P)$ where $\mathbf{C} = \{C_1, \ldots, C_r\}$ satisfy Postulates 1 through 6 (see [25]). Then it has the form*

$$G(\Omega, \mathbf{C}, P) = \sum_{j=1}^{r} u(C_j) P(C_j) \log \frac{1}{P(C_j)},$$

where $u(C_j) = \frac{\int_{C_j} u(\omega)\, dP(\omega)}{P(C_j)}$ and $u\colon \Omega \to \mathbb{R}$ is an integrable nonnegative function on the parameter space Ω.

We can see, in particular, that for the given question $\mathbf{C}$ its difficulty depends, besides the initial probability measure P, on the function $u(\omega)$ defined on the parameter space Ω. This function can be called – using parallels with thermodynamics described in [25] – *pseudo-temperature*. The the resulting question difficulty from Theorem 1 can be thought of as the amount of *pseudo-energy* associated with the question so that the more difficult questions are those with higher values of pseudo-energy. The expression $H(\Omega, \mathbf{C}, P) = \sum_{j=1}^{r} P(C_j) \log \frac{1}{P(C_j)}$ is referred to as *entropy* of question $\mathbf{C}$ and it coincides with Shannon entropy of the distribution $P(\mathbf{C}) \equiv (P(\mathbf{C}_1), \ldots, P(\mathbf{C}_r))$. It is easy to see that in case of constant pseudo-temperature the question pseudo-energy (difficulty) and its entropy are related by $G(\Omega, \mathbf{C}, P) = u(\Omega) H(\Omega, \mathbf{C}, P)$ which is identical to the relationship between thermal energy (heat) and entropy in thermodynamics (see [25] for more details).

4.2 Answers

Given a question $\mathbf{C}$ on Ω, an answer to $\mathbf{C}$ was defined in [24] to be a message $V(\mathbf{C})$ taking values in the set $\{s_1, \ldots, s_m\}$ such that the reception of the value s_k modifies (updates) the initial measure P on Ω to the measure $P^k \equiv P^{V(\mathbf{C})=s_k}$ such that $P^k_{C_j} = P_{C_j}$ for $k = 1, \ldots, m$ and $j = 1, \ldots, r\}$. The latter condition ensures that the answer $V(\mathbf{C})$ is indeed an answer to the question $\mathbf{C}$.

It follows from the above definition that, for $V(\mathbf{C})$ to be an answer to a multiple-choice question $\mathbf{C}$, it is necessary and sufficient for the updated measures P^k, $k = 1, \ldots, m$, to take the form

$$P^k = \sum_{j=1}^{r} p_{kj} P_{C_j}, \tag{6}$$

where p_{kj}, $k = 1, \ldots, m$, $j = 1, \ldots, r$ are nonnegative coefficients such that $\sum_{j=1}^{r} p_{kj} = 1$ for $k = 1, \ldots, m$.

For incomplete (free-response and mixed) questions, the expression (6) gets slightly modified to account for the set $\overline{\hat{C}} = \Omega \setminus \hat{C}$ and takes the form

$$P^k = \sum_{j=1}^{r} p_{kj} P_{C_j} + \bar{p}_k P_{\overline{\hat{C}}}, \tag{7}$$

where $\sum_{j=1}^{r} p_{kj} + \bar{p}_k = 1$. For pure free-response questions, one has to set $r = 1$ in (7).

The answer *depth* function $Y(\Omega, \mathbf{C}, P, V(\mathbf{C}))$ for the answer $V(\mathbf{C})$ to question $\mathbf{C}$ measures the amount of *pseudo-energy* that is conveyed by $V(\mathbf{C})$ in response to question $\mathbf{C}$. The general form of $Y(\Omega, \mathbf{C}, P, V(\mathbf{C}))$ can be established if certain (reasonable) requirements it has to satisfy are imposed. This was done in [24] where such requirements – called postulates – were discussed. The particular set of postulates used in [24] was chosen to impose *linearity* and *isotropy* conditions on the answer depth function. Under these conditions, the following result was obtained (formulated in [24] as a corollary).

Theorem 2 *The answer depth function $Y(\Omega, \mathbf{C}, P, V(\mathbf{C}))$ has the form*

$$Y(\Omega, \mathbf{C}, P, V(\mathbf{C})) = \sum_{k=1}^{m} \Pr(V(\mathbf{C}) = s_k) \frac{\sum_{j=1}^{r} u(C_j) P^k(C_j) \log \frac{P^k(C_j)}{P(C_j)}}{\sum_{j=1}^{r} P^k(C_j)},$$

where $P^k \equiv P^{V(\mathbf{C})=s_k}$ is the measure on Ω conditioned on reception of $V(\mathbf{C}) = s_k$ and $u(C_j) = \frac{1}{P(C_j)} \int_{C_j} u(\omega) dP(\omega)$ and the function $u\colon \Omega \to \mathbb{R}$ is the same function that is used in the question difficulty function $G(\Omega, \mathbf{C}, P)$.

It can be shown, in particular (see [24] for details) that if $V(\mathbf{C})$ is any answer to the question $\mathbf{C}$ then $Y(\Omega, \mathbf{C}, P, V(\mathbf{C})) \leq G(\Omega, \mathbf{C}, P)$ with equality if and only if the answer $V(\mathbf{C})$ is *perfect*, i.e. $P^j = P_{C_j}$ for $j = 1, \ldots, r$.

It is convenient to consider the class of imperfect answers for which the degree of imperfection is described by a single error probability α – the *quasi-perfect* answers [24]. For a quasi-perfect

answer $V_\alpha(\mathbf{C})$ to a (complete) question $\mathbf{C} = \{C_1, \ldots, C_r\}$, the coefficients p_{kj} have the form

$$p_{kj} = (1-\alpha)\delta_{k,j} + \alpha P(C_j), \tag{8}$$

for $k = 1, \ldots, r$ and $j = 1, \ldots, r$, and the updated measure P^k is simply

$$P^k = \alpha P + (1-\alpha)P_{C_k}. \tag{9}$$

for $k = 1, \ldots, r$.

4.3 Information Source

In addition to the knowledge of the probability measure P that embodies the original state of information available to the decision maker, an information source is assumed to be capable of answering questions of the form $\mathbf{C}$ discussed above. The answers $V(\mathbf{C})$ modify the original measure P on Ω. The questions differ from each other in the degree of difficulty that can be measured – under certain assumptions – by the question difficulty function whose general form is given in Theorem 1. The source's answers can be characterized by their depth $Y(\Omega, \mathbf{C}, P, V(\mathbf{C}))$ whose general form is established in Theorem 2. As was mentioned earlier, both the question difficulty and answer depth functions depend, besides the original and updated measures on Ω, on an integrable pseudo-temperature function u: $\Omega \to \mathbb{R}$ whose value at a point $\omega \in \Omega$ has the meaning of the "local difficulty" at that point. Therefore, if the function $u(\omega)$ is given then the difficulty of any question can be computed for any original measure P on Ω. On the other hand, in any real application, the function $u(\omega)$ cannot be known since it is not directly observable. What can be observed is the information source's actual performance: the proportion of correct answers. From that, the error probabilities can be estimated. This means that the function $u(\omega)$ has to be estimated from the knowledge of error probabilities exhibited by the information source in response to some particular questions. Informally speaking, the error probabilities tell us indirectly what questions are easy and which are hard for the information source. If we assume that the postulates discussed in [25] and [24] are valid (that is if the linear isotropic model is adequate) then the function $u(\omega)$ can be found that would reproduce – within estimation error – the observed (estimated) error probabilities. The estimated function $u(\omega)$, in turn, would allow for computation of difficulties of other questions that have not been given to the source before.

Let us recall the general assumptions that were made about the information source in [25] and [24].

1. Questions that can be given to the source have different degrees of detalization and difficulty.

2. A question's degree of difficulty is related to the question degree of detalization but in general does not coincide with it.

3. The quality of source's answers is directly related to the degree of difficulty of the corresponding questions.

4. The source has a finite capacity.

5. The source "tries equally hard" to answer any question it receives. Therefore, the source answers those questions well (i.e. with low error probabilities) whose difficulty does not exceed the source's capacity. As the difficulty exceeds the source's capacity the quality of its answers progressively degrades.

Assumptions 1 and 2 are subsumed by question difficulty postulates: the degree of detalization for the question $\mathbf{C} = \{C_1, \ldots, C_r\}$ can be identified with the number of subsets in the corresponding partition (in the "topological" sense) or with the expression $-\sum_{j=1}^{r} P(C_j) \log P(C_j)$ (in the "metric" sense) and its difficulty is given by $G(\Omega, \mathbf{C}, P)$. The latter is different from the "metric" degree of detalization by virtue of the presence of function $u(\omega)$ and reduces to it for the case of constant u.

Assumption 3 implies that the source answers questions in such a way that the quality of its answers measured by the answer depth function is in direct relation to the question difficulty – measured by the question difficulty function. More precisely, for the given information source, the answer depth has to be a function of the corresponding question difficulty. Assumptions 4 and 5 then imply that this function is non-decreasing and is bounded from above. We formalize these observations by adapting the following main hypothesis.

Hypothesis S1. For the given information source and any question $\mathbf{C}$, the corresponding answer depth is a function of the question difficulty:

$$Y(\Omega, \mathbf{C}, P, V(\mathbf{C})) = h(G(\Omega, \mathbf{C}, P)),$$

where $h(\cdot)$ is a non-decreasing function of its argument that's bounded from above.

The hypothesis S1 essentially states that the question difficulty and answer depth are exhaustive characterizations of the pseudo-energy content of questions and answers, respectively. If two

different questions have the same difficulty, the information source will answer them equally well, i.e. the depth of answers will be the same.

It is natural to call the particular form of function $h(\cdot)$ the *model of the source*. In practice, the overall form of $h(\cdot)$ has to be postulated. Then the values of parameters needed full specification of $h(\cdot)$ and the function $u(\omega)$ can be estimated from the observed performance of the source on sample questions.

5 Possible Source Models

As was mentioned above, the model of the source is described by a non-decreasing function $h(\cdot)$ where the role of the argument is played by the question difficulty $G(\cdot)$. The function $h(\cdot)$ should also be bounded from above if one assumes (as we do) that a source has a finite (effective) informational capacity. Let us now describe some possible models.

5.1 Simple Capacity Model

In this model, the information source is characterized by a single parameter that can be called the *pseudo-energy capacity* and denoted by Y_s. Under this model, the source can provide perfect answers to questions whose difficulty does not exceed Y_s and, for questions with difficulty exceeding Y_s, the error probabilities increase in such a way that the depth of the corresponding answer stays equal to Y_s. Put slightly differently, the information source provides answers whose depth is constant unless the question is too easy for the source in which case the depth of the answer is limited by the difficulty of the question. Formally speaking, the function $h(x)$ for this model takes the following form.

$$h(x) = \begin{cases} x & \text{if } x \leq Y_s \\ Y_s & \text{if } x > Y_s. \end{cases} \tag{10}$$

In reality, while one wouldn't expect a perfect fit of empirical data to 10, large deviations could indicate either inadequacy of the linear isotropic model of question difficulty or that of the capacity model 10 of the information source.

5.2 Modified Capacity Models

The main drawback of the simple capacity model described above is that the information source is postulated to provide perfect answers to questions whose difficulty is below the source's capacity. On the other hand, in many situations, it is reasonable to expect that a source will make some error answering even relatively simple questions. The modified capacity models' goal is to allow for finite error probabilities for answers to questions with difficulties below the source capacity. This model depends on more than one parameter: besides the capacity Y_s, there is also a parameter describing the approach by the function $h(\cdot)$ of its maximum value Y_s. The simplest of such models is the *linear* modified capacity model described by

$$h(x) = \begin{cases} bx & \text{if } x \leq \frac{Y_s}{b} \\ Y_s & \text{if } x > \frac{Y_s}{b}. \end{cases} \tag{11}$$

where $b \leq 1$ is the second parameter. Under this model, the source makes errors even on questions with difficulties below the capacity with error probabilities gradually increasing with question difficulties. Once the question difficulty exceeds the capacity of the source, the corresponding answer depth stays equal to the capacity Y_s.

The linear modified capacity model can be naturally generalized to the *polynomial modified capacity model* in which the function $h(\cdot)$ approaches its maximum value according to a polynomial law. To describe it, let $p_q(x) = a_0 + a_1 x + \ldots + a_q x^q$ be an order q polynomial and let x_q^* be the smallest positive root of the equation $p_q(x) - Y_s = 0$. Then the polynomial modified capacity model has the form

$$h(x) = \begin{cases} p_q(x) & \text{if } x \leq x_q^* \\ Y_s & \text{if } x > x_q^*. \end{cases} \tag{12}$$

Demanding that $h(0) = 0$ and $h(x) \leq x$ for all $x \geq 0$ leads to $a_0 = 0$ and $0 \leq a_1 \leq 1$. For $q = 2$, the polynomial modified capacity model (12) reduces to the *quadratic modified capacity model* that is most conveniently written in the form

$$h(x) = \begin{cases} bx - \frac{\gamma}{Y_s} x^2 & \text{if } x \leq G_2 \\ Y_s & \text{if } x > G_2, \end{cases} \tag{13}$$

where $0 < b \leq 1$ and (assuming $\gamma > 0$ so that $h(x)$ is concave) $\gamma \leq \frac{b}{4}$; $G_2 = \frac{b - \sqrt{b^2 - 4\gamma}}{2\gamma} Y_s$. In this model Y_s has the meaning of the source capacity and the coefficients b and γ are pure numbers

(i.e. their numerical values do not depend on the choice of units of pseudo-temperature $u(\cdot)$ and capacity Y_s).

Another simple model that belongs to the class of modified capacity models is the *exponential modified capacity model*

$$h(x) = Y_s(1 - e^{-\frac{\theta}{Y_s}x}) \tag{14}$$

that depends on two parameters: capacity Y_s and $0 < \theta \leq 1$ that controls the speed with which the function $h(x)$ approaches its upper bound Y_s. The coefficient θ is a pure number in the sense described above. One of the advantages of the exponential model (14) is that it is described by a single analytical function that allows the corresponding estimation problem that is discussed in the next section to avoid binary variables.

6 Estimation of Model Parameters and Function $u(\omega)$

First, let us note that both question difficulty and answer depth functions are linear in $u(\omega)$ and therefore multiplying $u(\omega)$ by any constant would result in both difficulty and depth being multiplied by the same constant without changing any of the coefficients p_{kj}, $k = 1, \ldots, m$, $j = 1, \ldots, r$ and therefore answer error probabilities. This means that the function $u(\omega)$ is really defined up to a single multiplicative constant the choice of which is equivalent to a choice of units in which $u(\omega)$ (and the difficulty/depth functions) are measured. We use two different conventions that turn out to be convenient.

- The normalized $u(\cdot)$ convention in which $\int_\Omega u(\omega)dP(\omega) = 1$ for every information source. This convention is convenient because if $u(\omega) \equiv 1$ the difficulty of question $\mathbf{C}$ reduces to Shannon entropy of the distribution $P(\mathbf{C}) = (P(C_1, \ldots, P(C_r))$.

- The unit source capacity convention in which the units of $u(\omega)$ are chosen in such a way that, for each information source, the capacity is unity: $Y_s = 1$. This convention is especially convenient for comparing different information sources to each other. Indeed, in this case, functions $u(\omega)$ for any two sources can be directly compared to each other showing clearly the relative degree of "expertise" of each source in various regions of Ω and also giving a sense of "absolute" quality of each source.

If the function $u(\omega)$ is known Theorem 1 gives – for the given measure P – the difficulty of any question $\mathbf{C}$. Then for any answer $V(\mathbf{C})$ to $\mathbf{C}$ the knowledge of updated measures P^k allows one to find the depth of $V(\mathbf{C})$. On the other hand, a given source model $Y = h(G)$ lets one *predict* the depth of the source's answer to any question before measures P^k can be estimated. Thus in order to be able to predict the depth of source's answer to various questions – and hence possibly solve the problem (3) – one needs to know (i) the function $u(\omega)$ and (ii) the source model described by the function $h(\cdot)$. Since these functions cannot be directly measured or observed, the only way to find these two functions in any realistic application is to estimate them from the source's performance on a certain set of sample questions.

Let $\mathbf{D} = \{D_1, \ldots, D_{N_d}\}$ be a partition of Ω to be used for discretizing the weight function $u(\omega)$: we assume that $u(\omega)$ takes a constant value equal to u_i on subset D_i. Let $w_i = P(D_i)$ and let $\mathcal{N}_i \subset \{1, \ldots, N_d\}$ be the index set of the subsets in $\mathbf{D}$ that are immediate neighbors of (i.e. have a common boundary with) subset D_i. We assume that the partition $\mathbf{D}$ is sufficiently fine so that any partition $\mathbf{C}$ used for estimating $u(\omega)$ can be considered a coarsening of $\mathbf{D}$.

Further, let $\mathbf{C}_1, \ldots, \mathbf{C}_K$ be a set questions that the source has answered and its answers have been compared with actual outcomes in Ω. Let us denote by $G_1, \ldots, G_K$ the difficulties of these questions and let $Y_1, \ldots, Y_K$ be the corresponding answer depth values that were computed using the estimated error probabilities. For the sake of simplicity, we assume that the answers of the source are quasi-perfect (see (8) and (9) for the form of coefficients p_{kj} and updated measures P^k) with the corresponding (estimated) error probabilities being equal to $\alpha_1, \ldots, \alpha_K$, respectively.

Let us denote $z_i = |Y_i - h(G_i)|$, $i = 1, \ldots, K$ where the function $h(\cdot)$ is given by the suitablei information source model. The quantities z_i measure the absolute values of deviations of the empirical data from the chosen source model, vanishing values of all variables z_i corresponding to a perfect fit. In addition to minimizing the sum of the deviations (i.e. maximizing the fit), it makes sense to demand that the quantities u_j, $j = 1, \ldots, N_d$, describe a reasonably smooth function $u(\omega)$. This can be achieved, for instance, by putting an upper bound on the gradient of $u(\omega)$ or, equivalently, by putting a corresponding term in the objective function. To make it more precise, let $N(\mathbf{D})$ be the set of neighbors in the partition $\mathbf{D}$ (i.e. $N(\mathbf{D}) = \{(i,j) : j \in \mathcal{N}_i, i = 1, \ldots N_d\}$) and let U be the desired upper bound on the difference of two values of u on neighboring sets of partition $\mathbf{D}$. Then if the capacity model $h(\cdot)$ is postulated, the following formulation of the estimation problem for the function $u(\omega)$ and the parameters of model $h(\cdot)$ is obtained.

$$
\begin{aligned}
\text{minimize} \quad & \sum_{i=1}^{K} z_i + \lambda U \\
\text{subject to} \quad & Y_i - h(G_i) \le z_i,\ i = 1, \ldots, K \\
& h(G_i) - Y_i \le z_i,\ i = 1, \ldots, K \\
& u_j - u_k \le U,\ (j,k) \in N(\mathbf{D}) \\
& u_k - u_j \le U,\ (j,k) \in N(\mathbf{D})
\end{aligned}
\tag{15}
$$

The decision variables in (15), besides z_i, are u_j, $j = 1, \ldots, N_d$ and the parameters of function $h(\cdot)$. The parameter λ controls the trade-off between the objective of maximizing the fit and that of maximizing smoothness of $u(\omega)$ (understood as minimizing the maximum gradient of $u(\omega)$). The difficulties G_i, $i = 1, \ldots, K$ are expressed via the decision variables as follows.

$$
G_i = -\sum_{j=1}^{r_i} \log P(C_j) \sum_{\{l: D_l \subset C_j\}} u_l w_l \tag{16}
$$

For the values of the depth function for the corresponding answers, let us assume, for simplicity that the answers are quasi-perfect implying that their errors can be characterized with a single probability α_i, $i = 1, \ldots, K$. Then the depth Y_i can be written as

$$
\begin{aligned}
Y_i = & \sum_{j=1}^{r_i} (1 - \alpha_i + \alpha_i P(C_j)) \log \frac{1 - \alpha_i + \alpha_i P(C_j)}{P(C_j)} \sum_{\{l: D_l \subset C_j\}} u_l w_l \\
& + \alpha_i \log \alpha_i \sum_{j=1}^{r_i} P(C_j) \left(1 - \sum_{\{l: D_l \subset C_j\}} u_l w_l \right).
\end{aligned}
\tag{17}
$$

Note that in general, (15) is a potentially complex nonlinear optimization problem where nonlinearity is introduced by the function $h(\cdot)$. For the case of the simple capacity model the problem (15) can be written as

$$
\begin{aligned}
\text{minimize} \quad & \sum_{i=1}^{K} z_i + \lambda U \\
\text{subject to} \quad & Y_i - Y_s \le z_i + M y_i,\ i = 1, \ldots, K \\
& Y_s - Y_i \le z_i + M y_i,\ i = 1, \ldots, K \\
& G_i - Y_i \le z_i + M(1 - y_i),\ i = 1, \ldots, K \\
& u_j - u_k \le U,\ (j,k) \in N(\mathbf{D}) \\
& u_k - u_j \le U,\ (j,k) \in N(\mathbf{D}) \\
& y_i \in \{0, 1\},\ i = 1, \ldots, K
\end{aligned}
\tag{18}
$$

In this formulation, M is a large number, y_i, $i = 1, \ldots K$ are auxiliary binary variables. The main decision variables in the formulation (18) are the values u_j, $j = 1, \ldots, N_d$ and the capacity value Y_s. Since both (16) and (17) are linear in the variables u_l, the optimization problem (18) is mixed-linear with K binary variables. Therefore, it can at least be solved efficiently for moderate values K of sample questions used for estimating model parameter Y_s and the (discretized) function $u(\omega)$.

The formulation (18) can be modified easily from the simple to the modified capacity model. The resulting formulation is as follows.

$$\begin{aligned}
\text{minimize} \quad & \sum_{i=1}^{K} z_i + \lambda U \\
\text{subject to} \quad & Y_i - Y_s \leq z_i + M y_i, \ i = 1, \ldots, K \\
& Y_s - Y_i \leq z_i + M y_i, \ i = 1, \ldots, K \\
& bG_i - Y_i \leq z_i + M(1 - y_i), \ i = 1, \ldots, K \\
& u_j - u_k \leq U, \ (j,k) \in N(\mathbf{D}) \\
& u_k - u_j \leq U, \ (j,k) \in N(\mathbf{D}) \\
& y_i \in \{0, 1\}, \ i = 1, \ldots, K
\end{aligned} \tag{19}$$

The additional decision variable in (19) is $b \leq 1$. The values G_i and Y_i, $i = 1, \ldots K$ are given by expressions (16) and (17), respectively. The formulation (19), just like (18), is a mixed-linear optimization problem with K binary variables and thus can at least be solved efficiently for moderate values of the number K of sample questions.

The formulation for the quadratic modified capacity model (13) can be easily obtained from (19) by replacing the constraints $bG_i - Y_i \leq z_i + M(1 - y_i)$, $i = 1, \ldots, K$ with $bG_i + cG_i^2 - Y_i \leq z_i + M(1 - y_i)$, $i = 1, \ldots, K$. Recalling that G_i is a linear function of the decision variables u_l, we see that the resulting problem is that of quadratic optimization with K binary variables that enter the formulation in a linear fashion. Even though such problems can't in general be solved as efficiently as mixed-linear optimization problems of equal size, they can still be solved to optimality for moderate values of parameters K and N_d.

As mentioned earlier, the exponential capacity model has the advantage over other models discussed here in that it obviates the need for binary variables even though it becomes severely

nonlinear:

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^{K} z_i + \lambda U \\ \text{subject to} \quad & Y_i - Y_s(1 - e^{-\theta G_i}) \leq z_i,\ i = 1, \ldots, K \\ & Y_s(1 - e^{-\theta G_i}) - Y_i \leq z_i,\ i = 1, \ldots, K \\ & u_j - u_k \leq U,\ (j,k) \in N(\mathbf{D}) \\ & u_k - u_j \leq U,\ (j,k) \in N(\mathbf{D}) \end{aligned} \tag{20}$$

Besides the quantities $z_i,\ i = 1, \ldots, K,\ u_l,\ l = 1, \ldots, N_d$ and the source capacity Y_s, another decision variable is the parameter $0 < \theta \leq \frac{1}{Y_s}$.

It is worth noting that in estimation of the pseudo-temperature function and model parameters the error probabilities are themselves estimated values. That introduces obvious imprecision in estimation of pseudo-temperature and source model parameters. In fact, one can think of the procedure described in this section as similar to point estimation of parameters in classical statistics. For more information about the pseudo-temperature function, confidence intervals would be needed. The width of such confidence intervals would obviously depend on the precision with which error probabilities are known and therefore on the sample size used in error probability estimation. Practically, such confidence intervals may turn out to be sufficiently wide to effectively invalidate precise estimation of the shape of pseudo-temperature function. The practical approach instead could be that of the hypothesis testing type: a null (default) hypothesis about the shape of the pseudo-temperature function would be stated (i.e. that the pseudo-temperature is constant or linear) and then tested using standard statistical methods.

Just like in probability estimation, expert opinion can be used for estimating pseudo-temperature function. Since pseudo-temperature admits a simple intuitive interpretation (as local "degree of difficulty") experts should find it easy enough to give useful estimates of pseudo-temperature. If, in addition, some data about observed source performance is available, it can be used in conjunction with expert estimates, for instance, by using these estimates as a null hypothesis and using observed data for the purpose of testing it.

7 Examples

To illustrate the process of estimation of the pseudo-temperature $u(\omega)$ and source model parameters, consider an example in which $\Omega = [0,1]^2 \subset \mathbb{R}^2$, and the measure P is uniform continuous on Ω. Consider the set of sample (complete) questions illustrated in Fig. 3. Our goal is, given the error parameters α_i for quasi-perfect answer $V_{\alpha_i}(\mathbf{C}_i)$ to question $\mathbf{C}_i$, $i = 1, \ldots, 10$, estimate the function $u(\omega)$ and the parameter(s) of the chosen information source model.

We adapt the modified linear source model and use formulation (19) to estimate $u(\omega)$, and parameters Y_s and b of the model. We do this for different values of error probabilities.

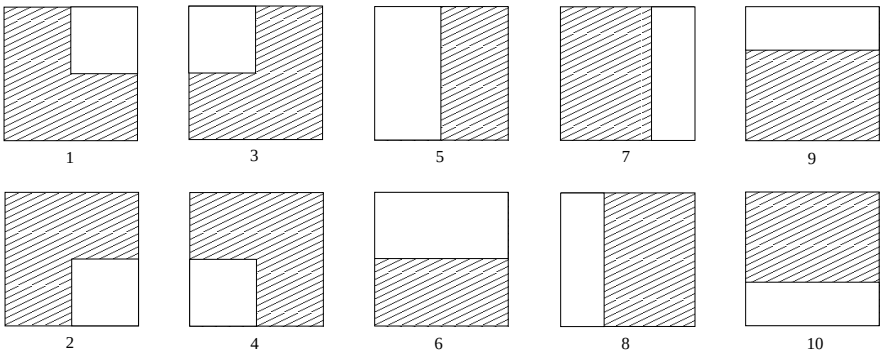

Figure 3: Sample questions.

First consider data shown in Table 1. In this and following tables, the first column contains the index i of question $\mathbf{C}_i$ from Fig. 3, the second column shows the corresponding error probability α_i, and the last two columns contain the question difficulty $G(\Omega, \mathbf{C}_i, P)$ and answer depth $Y(\Omega, \mathbf{C}_i, P, V_{\alpha_i}(\mathbf{C}_i))$, respectively, obtained from the estimated values of $u(\omega)$ and parameters of the source model. In the lower part of Table 1, the resulting value of the objective of problem (19) along with the estimated values of parameters Y_s and b are shown.

The error probability values shown in Table 1 result in a perfect fit ($z = 0$) with the estimated pseudo-temperature function $u(\omega)$ (shown in Fig. 4). We can see that the resulting pseudo-temperature function increases for the larger values of coordinates ω_1 and ω_2 on Ω reflecting the fact that, for instance $\alpha_1 > \alpha_4$, implying that question $\mathbf{C}_1$ has higher difficulty (larger value of pseudo-energy) than $\mathbf{C}_4$ in spite of these two questions having same value of entropy. This means that the smaller measure subset in $\mathbf{C}_1$ has to have higher pseudo-temperature which we indeed see. It is also worth noting that questions $\mathbf{C}_5$ and $\mathbf{C}_6$ were answered with equal accuracy suggesting that

Table 1: Sample question error probabilities, fitted values of the difficulty and depth functions, and estimated model parameter values for the modified linear model when perfect fit is possible.

i	α_i	$G(\Omega, \mathbf{C}_i, P)$	$Y(\Omega, \mathbf{C}_i, P, V_{\alpha_i}(\mathbf{C}_i))$
1	0.265	1.106	0.516
2	0.143	0.803	0.516
3	0.143	0.803	0.516
4	0.077	0.533	0.404
5	0.210	1.000	0.516
6	0.210	1.000	0.516
7	0.253	1.102	0.516
8	0.116	0.761	0.516
9	0.253	1.102	0.516
10	0.116	0.761	0.516

$\sum_{i=1}^{N_d} z_i = 0$; $U = 0.13$; $Y_s = 0.52$; $b = 0.76$.

these questions are of equal difficulty. This in fact is a necessary condition for a perfect fit within the ideal gas question difficulty model since in this model any complete question with all subsets of equal measure would have the same difficulty (pseudo-energy) regardless of the pseudo-temperature function form.

Consider now data shown in Table 2. The resulting pseudo-temperature $u(\omega)$ is shown in Fig. 5. We see that in this case the perfect fit could not be achieved by any pseudo-temperature function, in particular because questions $\mathbf{C}_5$ and $\mathbf{C}_6$ were answered with slightly different accuracy whereas these two questions necessarily have equal pseudo-energy content (equal difficulty) within the ideal gas question difficulty model.

Now, consider the data shown in Table 3. As can be seen from Fig. 6, the fit that could be achieved to the ideal gas question difficulty model (with the linear modified information source model) is relatively (at least compared to the previous example) poor, possibly indicating that the ideal gas model may not be adequate in this case and that a different model (for example, anisotropic – to be able to model different pseudo-energy content of questions $\mathbf{C}_5$ and $\mathbf{C}_6$) may be needed.

Let us now turn to comparing different sources. Suppose $\Omega = [0, 1]$ with P being a uniform

Table 2: Sample question error probabilities, fitted values of the difficulty and depth functions, and estimated model parameter values for the modified linear model when perfect fit is not possible, with small misfit.

i	α_i	$G(\Omega, \mathbf{C}_i, P)$	$Y(\Omega, \mathbf{C}_i, P, V_{\alpha_i}(\mathbf{C}_i))$
1	0.238	1.057	0.531
2	0.157	0.856	0.531
3	0.129	0.794	0.531
4	0.084	0.538	0.399
5	0.189	1.000	0.549
6	0.230	1.000	0.484
7	0.227	1.055	0.531
8	0.127	0.806	0.531
9	0.278	1.200	0.525
10	0.127	0.806	0.531

$\sum_i^{N_d} z_i = 0.07$; $U = 0.43$; $Y_s = 0.53$; $b = 0.74$.

Table 3: Sample question error probabilities, fitted values of the difficulty and depth functions, and estimated model parameter values for the modified linear model when perfect fit is not possible, with larger misfit.

i	α_i	$G(\Omega, \mathbf{C}_i, P)$	$Y(\Omega, \mathbf{C}_i, P, V_{\alpha_i}(\mathbf{C}_i))$
1	0.371	0.418	0.118
2	0.086	0.488	0.358
3	0.200	0.589	0.312
4	0.107	1.750	1.281
5	0.126	1.000	0.661
6	0.293	1.000	0.399
7	0.354	0.585	0.180
8	0.162	1.320	0.812
9	0.354	0.590	0.182
10	0.162	1.219	0.746

$\sum_i^{N_d} z_i = 1.51$; $U = 0.56$; $Y_s = 1.28$; $b = 0.73$.

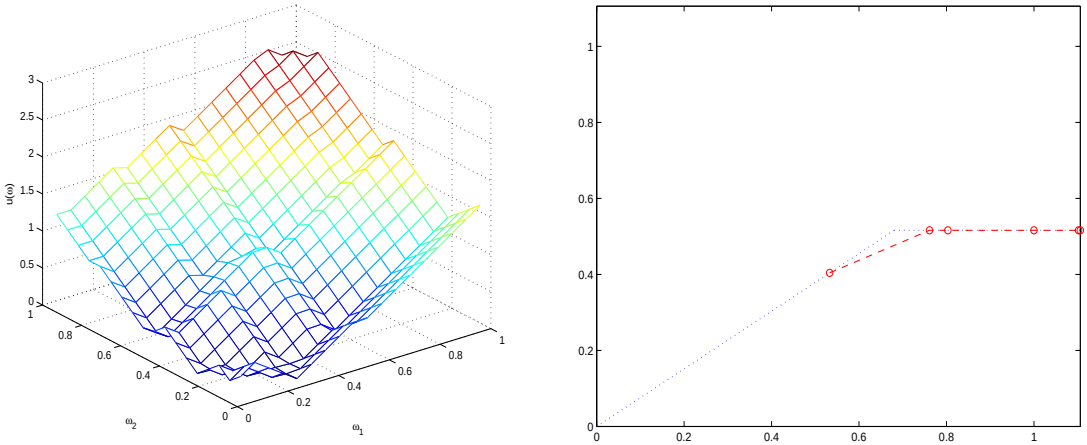

Figure 4: The estimated pseudo-temperature (left) and the fitted values of difficulty and depth (right) for the data of Table 1.

continuous measure on Ω. Let sample questions be as follows. $\mathbf{C}_1 = \{[0, 1/2], (1/2, 1]\}$, $\mathbf{C}_2 = \{[0, 1/3], (1/3, 1]\}$, $\mathbf{C}_3 = \{[0, 2/3], (2/3, 1]\}$, $\mathbf{C}_4 = \{[0, 1/4], (1/4, 1]\}$, $\mathbf{C}_5 = \{[0, 3/4], (3/4, 1]\}$. Let source 1 accuracy be described by error probabilities (assuming quasi-perfect answers as before) shown in Table 4. Then, using the modified capacity model and formulation (19), we can estimate the pseudo-temperature function $u(\cdot)$ and the model parameters Y_s and b. The results – as well as fitted values of the question difficulty and answer depth – are shown in Table 4.

Table 5 shows error probabilities achieved on the same set of sample questions by a different source 2, along with the resulting fitted values of difficulty and depth functions and the estimated model parameter values. Looking at Tables 4 and 5 we can see, for example, that source 1 shows better overall performance on all questions, but there exist questions (question 5, for instance) that appear to be easier for source 2. Indeed, the estimated pseudo-temperature functions shown in Fig. 7 (in the unit source capacity convention) clearly demonstrate that the overall pseudo-temperature is significantly higher for source 2 thus making the majority of sample questions more difficult for it (which is reflected in higher error probabilities). On the other hand, while the pseudo-temperature function for source 1 is (mostly) increasing on the interval $[0, 1]$, it is a decreasing function on the same interval for source 2. In particular, there exist regions of $\Omega = [0, 1]$ where the pseudo-temperature for source 2 is lower than that for source 1. This means that some questions can be easier for source 2, question 5 from the sample set being an example.

Table 4: Sample question error probabilities, fitted values of the difficulty and depth functions, estimated model parameter values for the modified linear model, for information source 1.

i	α_i	$G(\Omega, \mathbf{C}_i, P)$	$Y(\Omega, \mathbf{C}_i, P, V_{\alpha_i}(\mathbf{C}_i))$
1	0.090	1.000	0.735
2	0.070	0.678	0.525
3	0.153	1.174	0.735
4	0.070	0.528	0.408
5	0.146	1.131	0.735

$\sum_i^{N_d} z_i = 0.09$; $U = 0.54$; $Y_s = 0.74$; $b = 0.77$.

Table 5: Sample question error probabilities, fitted values of the difficulty and depth functions, estimated model parameter values for the modified linear model, for information source 2.

i	α_i	$G(\Omega, \mathbf{C}_i, P)$	$Y(\Omega, \mathbf{C}_i, P, V_{\alpha_i}(\mathbf{C}_i))$
1	0.300	0.933	0.386
2	0.350	1.000	0.331
3	0.170	0.415	0.229
4	0.350	1.115	0.386
5	0.080	0.585	0.434

$\sum_i^{N_d} z_i = 0.18$; $U = 0.56$; $Y_s = 0.39$; $b = 0.74$.

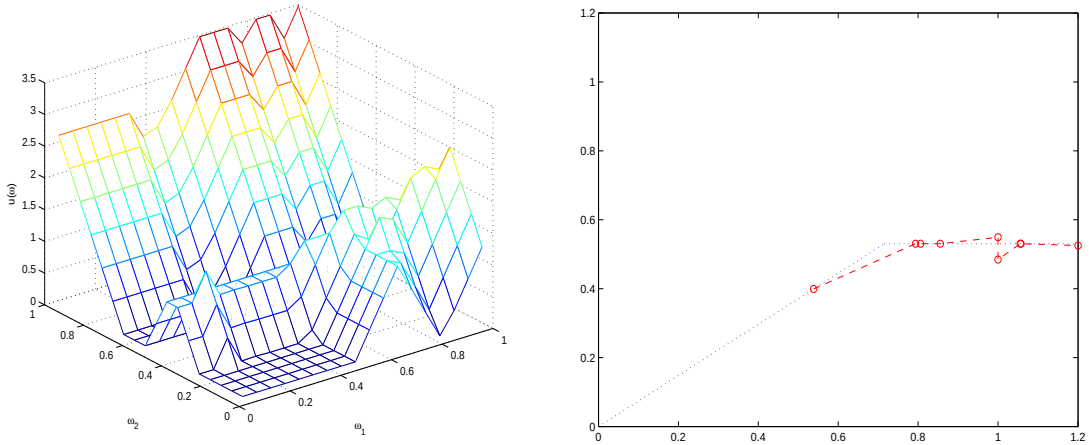

Figure 5: The estimated pseudo-temperature (left) and the fitted values of difficulty and depth (right) for the data of Table 2.

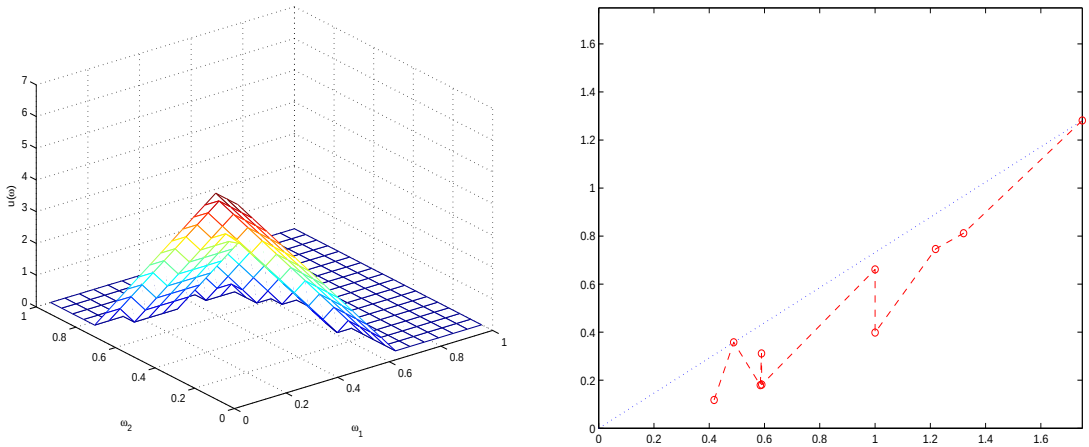

Figure 6: The estimated pseudo-temperature (left) and the fitted values of difficulty and depth (right) for the data of Table 3.

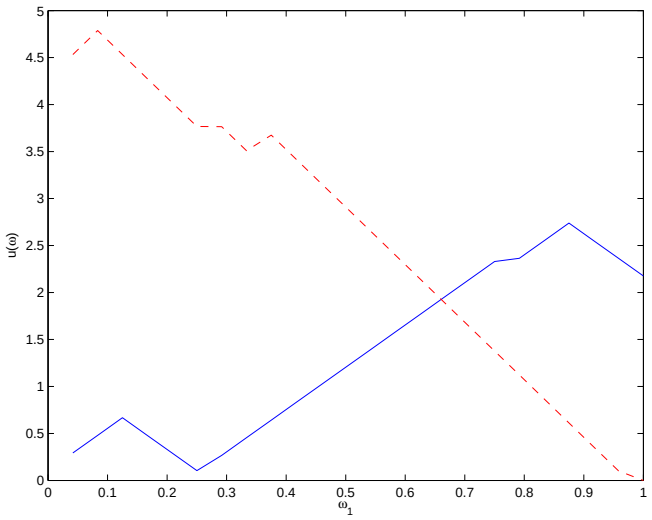

Figure 7: Estimated pseudo-temperature functions for information sources 1 and 2.

8 Conclusion

This article is the last in the series of three closely connected papers devoted to the development of a quantitative framework describing the process of additional information acquisition in decision making problems under uncertainty. The proposed framework is based on the assumption that the decision maker has access to one or more sources of information capable of answering questions concerning the problem parameter space, or, equivalently, the space of uncertain problem parameters (input data). A decision maker's questions and a source's respective answers were considered in [25] and [24], respectively. The question difficulty function introduced and studied in [25] can serve as a quantitative measure of the degree of difficulty of various questions for the given source. The main idea is that the knowledge of this function of the source allows the decision maker to predict the degree of accuracy of the source's possible answers to various questions and thus enables the decision maker to determine the particular question(s) that need to be asked to the given source in order to maximize the answer's impact on the solution quality for the given problem. The answer depth function studied in [24] provides a quantitative measure of the "amount of work" the source has to do in order to provide an answer of given accuracy to the question at hand. Roughly speaking, the main idea here is that the source would not be able to answer difficult question accurately because the answer depth required to make the answer accurate would exceed to source's capability. And it is the latter that is the main subject of the present paper.

The main goal of the present article is twofold: to study possible models of information sources and to propose methods for estimation of model parameters from the observed source's performance on sample questions. Information source models quantitatively express the idea that an information source can answer easy question more accurately than difficult ones. More precisely, the source's answer depth is limited just by question difficulty for questions that are easy enough and by the source's capability for more difficult questions. This simple and natural idea is quantified by the information source model that is a functional dependence of the answer depth on the question difficulty. It is easy to see that such a function has to be nondecreasing and has to approach a finite value for large values of the argument. In this paper, several such functions were proposed.

As was shown in [25] and [24], both the question difficulty and answer depth functions are described, besides appropriate probability measures, by a scalar function on the problem parameter space – termed pseudo-temperature in [25] using parallels with thermodynamics. In real applica-

tions, this function needs to be estimated along with source model parameters, from the observed source performance on a set of sample questions. In this paper, optimization based algorithms for estimating the pseudo-temperature function (using a suitable discretization of the parameter space) and the chosen source model parameters were proposed.

Finally, it is worth mentioning that the developments in [25], [24] and the present article were all based on the assumption that both the question difficulty and answer depth possess linearity and isotropy (on the problem parameter space) properties that – using parallels with thermodynamics – were referred to as the "ideal gas model". While this particular assumption leads to a concise and attractive form of the difficulty and depth functions, it is entirely possible that more general (i.e. anisotropic) models would be required for accurate description of performance of realistic information sources. Such generalizations will be the subject of future publications.

*Bibliography

[1] Bickel, E. 2007. Some comparisons among quadratic, spherical, and logarithmic scoring rules. *Decision Anal.* **4**(2) 49–65.

[2] Bordley, R. F. 2009. Combining the opinions of experts who partition events differently. *Decision Anal.* **6**(1) 38–46.

[3] Bordley, R. F. 2011. Using Bayes' rule to update an event's probabilities based on the outcomes of partially similar events. *Decision Anal.* **8**(2) 117–127.

[4] Brier, G. W. 1950. Verification of forecasts expressed in terms of probabilities. *Monthly Weather Rev.* **78**(1) 1–3.

[5] Chávez, M., J. Martinerie, M. Le Van Quyen. 2003. Statistical assessment of nonlinear causality: Application to epileptic EEG signals. *J. of Neurosci. Methods* **124**(2) 113–128.

[6] Clemen, R. 1987. Combining overlapping information. *Management Sci.* **33**(3) 373–380.

[7] Clemen, R., R. Winkler. 1999. Combining probability distributions from experts in risk analysis. *Risk Anal.* **19**(2) 187–203.

[8] Ellison, G., D. Fudenberg. 1993. Rules of thumb for social learning. *J. Political Econom.* **101**(4) 612–643.

[9] Fischer, A. J., A. J. Arnold, M. Gibbs. 1996. Information and the speed of innovation adoption. *Amer. J. Agr. Econ.* **78**(4) 1073–1081.

[10] Fisher, M. L., A. Raman. 1996. Reducing the cost of demand uncertainty through accurate response to early sales. *Oper. Res.* **44**(1) 87–99.

[11] Fox, C., R. Clemen. 2005. Subjective probability assessment in decision analysis: Partition dependence and bias toward the ignorance prior. *Management Sci.* **51**(9) 1417–1432.

[12] French, S. 1985. Group consensus probability distributions: A critical survey. *Bayesian Statist.* **2** 183–202.

[13] Genest, C., J. V. Zidek. 1986. Combining probability distributions: A critique and an annotated bibliography. *Statist. Sci.* **1** 114–148.

[14] Gneiting, T., A. Raftery. 2007. Strictly proper scoring rules, prediction, and estimation. *J. Amer. Statist. Assoc.* **102** 359–378.

[15] Good, I. J. 1952. Rational decisions. *J. Roy. Statist. Soc. Ser. B* **14**(1) 107–114.

[16] Jaynes, E. T. 1957. Information theory and statistical mechanics I. *Phys. Rev.* **106** 620–630.

[17] Jaynes, E. T. 1957. Information theory and statistical mechanics II. *Phys. Rev.* **108** 171–190.

[18] Jensen, R. 1988. Information cost and innovation adoption policies. *Management Sci.* **34**(2) 230–239.

[19] Johnstone, D. L., V. R. R. Jose, R. L. Winkler. 2011. Tailored scoring rules for probabilities. *Decision Anal.* **8**(4) 256–268.

[20] Katura, T., N. Tanaka, A. Obata, H. Sato, A. Maki. 2006. Quantitative evaluation of interrelations between spontaneous low-frequency oscillations in cerebral hemodynamics and systemic cardiovascular dynamics. *NeuroImage* **31**(4) 1592–1600.

[21] Kornish, L. J., R. L. Keeney. 2008. Repeated commit-or-defer decisions with a deadline: The influenza vaccine composition. *Oper. Res.* **56**(3) 527–541.

[22] McCardle, K. F. 1985. Information acquisition and the adoption of new technology. *Management Sci.* **31**(11) 1372–1389.

[23] Mokhov, I. I., D. A. Smirnov. 2006. El Niño-Southern Oscillation drives North Atlantic Oscillation as revealed with nonlinear techniques from climatic indices. *Geophys. Res. Lett.* **33**. L03708.

[24] Perevalov, E., D. Grace. 2011. Information acquisition for decision making: Answer depth. Available at `http://www.lehigh.edu/~dpg3/opt-info-acq.html`.

[25] Perevalov, E., D. Grace. 2011. Information acquisition for decision making: Question difficulty. Available at `http://www.lehigh.edu/~dpg3/opt-info-acq.html`.

[26] Predd, J. B., D. N. Osherson, S. R. Kulkarni, H. V. Poor. 2008. Aggregating probabilistic forecasts from incoherent and abstaining experts. *Decision Anal.* **5**(4) 177–189.

[27] Savage, L. J. 1971. Elicitation of personal probabilities and expectations. *J. Amer. Statist. Assoc.* **66** 783–801.

[28] Verdes, P. F. 2005. Assessing causality from multivariate time series. *Phys. Rev. E* **72**. 026222.

[29] Viola, P. A. 1995. Alignment by maximization of mutual information. A.I. Technical Report 1548, Massachusetts Institute of Technology.

[30] Winkler, R. L. 1996. Scoring rules and the evaluation of probabilities. *Test* **5** 1–60.