Information Acquisition for Decision Making: Answer Depth

David Grace
Lehigh University

Eugene Perevalov
Lehigh University

Report: 12T-002

Information Acquisition for Decision Making: Answer Depth

Draft

Abstract

The problem of optimal decision making in environments characterized by uncertainty and in the presence of information sources is considered in a general setting. This brings about the need for a quantitative framework for the description of information exchange between the decision maker and information sources. The difficulty of decision maker's questions was considered previously in a companion paper. Here, information sources' answers are investigated. In particular, the concept of answer depth is introduced that quantifies the amount of suitably defined effort required to provide an answer of a given accuracy. The overall form of the answer depth function is derived by demanding that it satisfy a particular set of postulates expressing, besides some reasonable consistency conditions, the linearity and isotropy properties. The latter properties justify calling the resulting information exchange model the "ideal gas model" making use of potentially fruitful parallels with classical thermodynamics.

1 Introduction

While extensive research has been performed on methods of decision making under incomplete information, little has been done about possible general approaches to optimizing the process of additional information acquisition in situations when extra information is in principle available. This article is part of an attempt at developing such methodology that was initiated in [21]. At the core of the proposed approach there are two main entities: a decision making problem (for instance, admitting a stochastic optimization formulation) and a source of information that is capable of answering questions concerning input data for the problem. Then the task of optimizing the additional information acquisition process can be formulated as that of finding the optimal question(s) such that the respective answer(s) would lead to the largest possible improvement in the solution quality for the problem. In other words, the optimal additional information acquisition task can be formulated as a problem of finding an optimal "alignment" between the problem and the information source that can be achieved by determining the particular question(s) capable of extracting from the source the largest amount of information relevant to the problem. Such optimal question(s) would obviously depend on both the problem and the information source.

Achieving the stated goal obviously requires developing an appropriate quantitative theory of information sources, questions and answers. The theory of questions was described in [21]. The present article is devoted to the study of answers.

1.1 Related Work

This idea of using additional information to improve the decision quality has been studied in the area of statistical decision making. One can mention applications to innovation adoption [19, 16], fashion decisions [9] and vaccine composition decisions for flu immunization [18]. 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) for accounting for the actual amount of information contained in the received observations [8, 6]. 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. The approach developed in this and the companion papers [21, 22] is different in that it allows one to optimize the *content* of the acquired information and it considers explicit models of information sources.

The problem of optimal usage of information obtained from experts has been addressed mostly in the form of updating the decision maker's beliefs given a probability assessment from multiple experts [10, 11, 4, 5] and, in particular, optimal combining of expert opinions, including experts with incoherent and missing outputs [23]. In particular, investigations on combining information of experts that partition the event differently [1] and on rules of updating probabilities based on outcomes of partially similar events [2] are close in spirit to the approach developed here in that they deal with different types of information. The emphasis of the proposed approach is on *optimizing* the particular type of information and on the explicit consideration of the dependence of the optimal information on both the expert and the decision making problem.

This paper uses an axiomatic approach to determine the overall form of the answer depth function. The latter, together with the related concept of question difficulty studied in [21] can be thought of as a logical development of the entropy concept of information theory. The axiomatic approach was used in [7] to derive the most general form of the (Shannon) entropy function. [24] used a different set of axioms to find the one-parameter family of functions (known as Rényi entropies) that included standard entropy as a special case. [12] introduced the concept of structural entropy and used for classification purposes. The Havrda-Charvat entropy was derived by axiomatic means in [25] where axiomatization of partition entropy was discussed on rather general grounds (see also [13]). It was shown in [25] that Shannon entropy, Havrda-Charvat entropy and Gini index all can be obtained as particular cases of general partition entropy that satisfies a system of reasonable axioms.

It should also be mentioned that this paper is part of the overall effort to bring Information Theory into closer contact with Operations Research in general and decision making and optimization in particular. Besides fundamental advances in communications, the list of successful applications of these concepts includes (but by no means is limited to) a simple derivation of statistical physics laws [14, 15], new algorithms in computer vision [27], new methods of analysis in climatology [20, 26], physiology [17] and neurophysiology [3]. The concept of pseudo-energy introduced in [21] and used in the present article builds on that of entropy and provides the foundation for the quantitative description of knowledge possessed by various information sources.

1.2 Outline

The rest of the article is organized as follows. In section 2, we briefly explain the main motivation for the proposed framework. In section 3, necessary preliminaries are given. In section 4, the three main information-related components of the proposed framework are briefly described and put in the overall context. In section 5, the overall form of the answer depth function is derived from a set of plausible postulates. Section 6 describes the main relationship between question difficulty and answer depth for main types of possible questions. In section 7, a special class of answers – the quasi-perfect answers – is discussed. Section 8 is devoted to relationship between different questions and, in particular, the relative depth of an answer to one question with respect to another question is introduced. Section 9 contains simple numerical example illustrating concepts and results discussed earlier in the article. Finally, section 10 gives a short summary of main results.

2 Motivation: Decision Making Under Uncertainty

The main motivation for the proposed framework was discussed in the companion paper [21]. We briefly recap it here so that the present paper can be read independently. If uncertainty is present in a decision making problem it can de described as a certain base space Ω that contains all possible sets of input data for the problem. The problem itself can be formulated as optimization with respect to a suitably chosen criterion. One such 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, with respect to a probability measure P that, in this formulation, describes the information available to the decision maker:

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

When uncertainty is present, 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. In the case where the optimization criterion is that of the expected objective function, as in (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 the random outcomes on Ω. If we denote the set of all possible questions that can be directed towards the source of information by $\mho$ and its answer to a particular question $Q \in \mho$ – 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.

Then 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 entail the minimum possible loss:

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

3 Preliminaries

As was already stated in the companion paper [21], 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 then the conditional measure P_C on Ω is defined by

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

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 $\mathbf{C}''$.

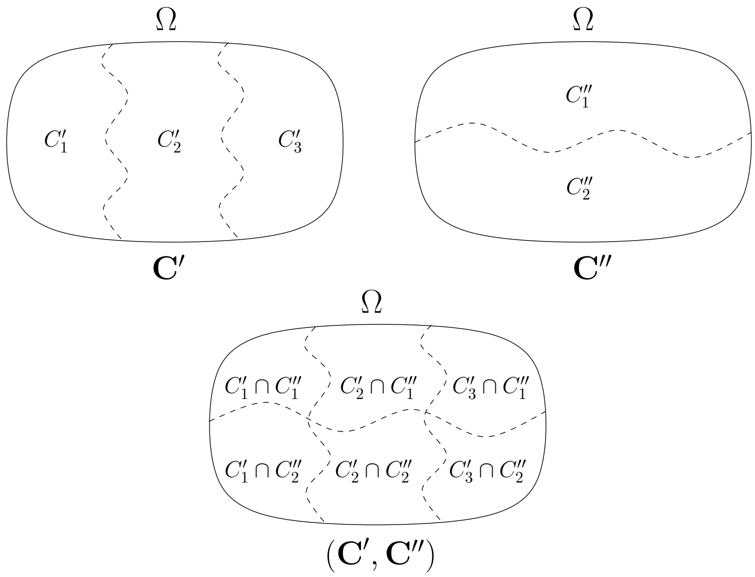

Figure 1: Two partitions of Ω and the corresponding joint partition.

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).

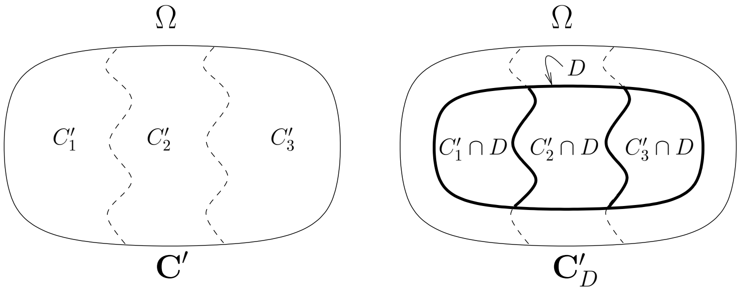

Figure 2: Partition $\mathbf{C}_D$ of set $D \subset \Omega$ induced by a partition $\mathbf{C}$ of Ω.

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 $\hat{C} = \Omega$.

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.

For an arbitrary complete partition $\mathbf{C} = \{C_1, \ldots, C_r\}$, the following decomposition of the measure P into the corresponding conditional measures is valid.

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

4 Overall Framework: Main Components

The overall setup is discussed in the companion paper [21]. The main components involved are the information source, questions concerning random outcomes on Ω and its answers to these questions. The article [21] is mainly devoted to the study of questions, while the main objective of the present article is an investigation of possible answers and their *depth* functions. To make the presentation self-contained, we briefly discuss the other components as well.

4.1 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 present that is capable of answering questions concerning random outcomes on Ω. These answers modify the original measure on Ω. The following rather general assumptions about the source are made that are explored in more detail in the companion paper [22].

- The source has a finite capacity (appropriately defined).
- Questions that can be given to the source have, in general, different degrees of detalization (elaborateness) and difficulty.
- A question's degree of difficulty is related to the question degree of detalization but in general does not coincide with it.
- The quality of source's answers is directly related to the degree of difficulty of the corresponding questions.
- 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.

4.2 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 [21] 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 [21] where a particular system of postulates that embodied *linearity* and *isotropy* properties of the difficulty function was proposed. The main theorem proved in [21] 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 [21]). Then it has the form*

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

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.

In particular, the difficulty of the given question $\mathbf{C}$ depends on, besides the initial probability measure P, the function $u(\cdot)$ defined on the parameter space Ω. This function may be called the *pseudo-temperature* using parallels with thermodynamics (see [21] for more details). The question difficulty then can be interpreted as the amount of *pseudo-energy* associated with question $\mathbf{C}$.

4.3 Answers

Given a question $\mathbf{C}$, a source is capable of providing an answer. Since any information in this context can be represented by some measure on P, it is reasonable to think of an answer to question $\mathbf{C}$ as a message the reception of which implies certain changes in the initial measure P. In an extreme case, a message can change the original measure to a measure supported at a single element of Ω – this describes a complete resolution of the initial uncertainty and to the best possible answer to the corresponding (exhaustive) question.

Thus, given a question $\mathbf{C}$, it makes sense to define an answer $V(\mathbf{C})$ to it as a message that can take values in the set $\{s_1, s_2, \ldots, s_m\}$, where s_k, $k = 1, \ldots, m$ is some symbolic string the length of which does not play an important role in the present context. Then, the conditional measure $P^{V(\mathbf{C})=s_k} \equiv P^k$ is in general different from the original measure P following a reception of the value s_k of message $V(\mathbf{C})$. Additionally, care has to be taken to ensure that the answer $V(\mathbf{C})$ is indeed an answer to the specific question $\mathbf{C}$ and not some other question. To achieve this we can require that a reception of $V(\mathbf{C})$ leave the relative likelihood of the elements inside every subset in $\mathbf{C}$ unchanged. Therefore probability is only "redistributed" between the members of $\mathbf{C}$. This way, an answer can't provide more information than what was requested in the question. We arrive at the following definition.

Definition: An answer to the question $\mathbf{C} = \{C_1, \ldots, C_r\}$ is a message $V(\mathbf{C})$ that takes values in the set $\{s_1, s_2, \ldots, s_m\}$ and such that $P^k_{C_j} = P_{C_j}$ for all $k = 1, \ldots m$ and all $j = 1, \ldots, r$.

Following this definition, it is straightforward to show 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, $r = 1$.

While the function $G(\Omega, \mathbf{C}, P)$ measures difficulty of questions, it would be desirable to develop a measure of the amount of difficulty in $\mathbf{C}$ that is resolved by the answer $V(\mathbf{C})$. As mentioned earlier, the question difficulty can be interpreted as the amount of *pseudo-energy* associated with the question. Therefore, it is natural to think that a perfect answer would contain an amount of pseudo-energy equal to the amount in the question it answers. Any other answer would contain somewhat less pseudo-energy, as long as it is an answer to $\mathbf{C}$ and not some other – possibly more difficult – question.

In the following we denote the amount of pseudo-energy contained in the answer $V(\mathbf{C})$ – the *depth* of $V(\mathbf{C})$ – by $Y(\Omega, \mathbf{C}, P, V(\mathbf{C}))$ to emphasize its dependence on Ω and the initial measure P.

5 Answer Depth Function

In this section, our goal is to derive the general form of the answer depth function by imposing certain plausible requirements it has to satisfy. These requirements that we call Postulates are similar to those stated in Postulates Q1 through Q6 for questions (see [21]).

Information in $V(\mathbf{C})$ is conveyed by modifying the original measure P and it modifies P differently for each value of the message $V(\mathbf{C})$. Therefore, the depth function for the message $V(\mathbf{C})$ should be the weighted average of the conditional values of the depth:

$$Y(\Omega, \mathbf{C}, P, V(\mathbf{C})) = \sum_{k=1}^{m} \Pr(V(\mathbf{C}) = s_k) Y(\Omega, \mathbf{C}, P, P^k), \tag{8}$$

where P^k is the measure modified by the reception of $V(\mathbf{C}) = s_k$ and $Y(\Omega, \mathbf{C}, P, P^k)$ is the conditional depth that depends on the modified measure P^k.

We now impose reasonable requirements on conditional depth functions $Y(\Omega, \mathbf{C}, P, P^k)$ which are formulated as postulates as before.

The first such requirement is that the conditional depth should vanish if the measure is not modified at all, i.e. if $P^k = P$. On the other hand, if the modified measure assigns larger probabilities to *all* subsets in $\mathbf{C}$ (which can happen only for incomplete – free-response and mixed – questions) then the conditional depth should be strictly positive. This is the content of Postulate A1.

Postulate A1 (*Correct direction*). Let $\mathbf{C} = \{C_1, \ldots, C_r\}$ be any question. Then $Y(\Omega, \mathbf{C}, P, P) = 0$ if $P^k(C_j) = P(C_j)$ for all $j = 1, \ldots r$ and $Y(\Omega, \mathbf{C}, P, P) > 0$ if $P^k(C_j) > P(C_j)$ for all $j = 1, \ldots r$.

The second part of the postulate says that, for a free-response question for instance, if upon reception of the value s_k of $V(\mathbf{C})$ the set C has a higher probability than before, then the value s_k has a positive amount of pseudo-energy. For example, if the original question was *"What kind of fruit is it?"* with *"Pear"* being the correct answer then in case the answer sounds like *"It looks a lot like a pear"* or *"It's either a pear or an apple"*, such an answer is assigned positive pseudo-energy as it moves "in the right direction" towards the correct answer.

The next postulate parallels Postulate Q2 for questions (see [21]).

Postulate A2 (*Continuity*). The function $Y(\Omega, \mathbf{C}, P, P^k)$ is continuous in all parameters it may depend upon.

The next postulate follows from the notion that $V(\mathbf{C})$ is an answer to precisely the question $\mathbf{C}$ and therefore the depth of $V(\mathbf{C})$ cannot exceed the difficulty of $\mathbf{C}$. The property is easiest to formulate for free-response questions $C \subset \Omega$.

Postulate A3 (*Free-response complete answer*). Let C be a free-response question and suppose

$P^k(C) = 1$. Then

$$Y(\Omega, C, P, P^k) = G(\Omega, C, P).$$

This postulate expresses a simple desideratum that an exhaustive correct answer to a question should convey exactly the amount of information requested by the question. For instance, if the question is *"What fruit is it?"* with *"Apple"* as a correct answer then the answer *"Apple"* should carry all the information the question was asking for.

The next three postulates parallel Postulates Q3 through Q5 for questions (see [21]).

Postulate A4 (*Mixed question answer decomposition*) Let $\mathbf{C} = \{C_1, \ldots, C_r\}$ be a mixed question. Then

$$Y(\Omega, \mathbf{C}, P, P^k) = Y(\Omega, \hat{C}, P, P^k) + Y(\hat{C}, \mathbf{C}, P_{\hat{C}}, P^k_{\hat{C}}).$$

This postulate states that the amount of pseudo-energy contained in a particular answer (value of $V(\mathbf{C})$ to a mixed question $\mathbf{C}$ can be represented as a sum of two components: the pseudo-energy of the same answer to the (free-response) question $\hat{C}$ and that of the same answer assuming the free-response part has been answered correctly. For example, if the question $\mathbf{C}$ is *"What kind of fruit is it and is it red, green or yellow?"* then Postulate A4 says that the depth of any particular answer to $\mathbf{C}$ is equal to the sum of the depth of the same answer to the question *"What kind of fruit is it?"* and the depth of the same answer to the question *"Is this apple red, green o yellow?"* (assuming the fruit in question was indeed an apple).

Postulate A5 (*Mean value*). Let $\mathbf{C}$ and $\mathbf{C}'$ be two incomplete questions such that $\hat{C} \cap \hat{C}' = \emptyset$. Then

$$Y(\Omega, \mathbf{C} \cup \mathbf{C}', P, P^k) = \frac{P^k(\hat{C})Y(\Omega, \mathbf{C}, P, P^k) + P^k(\hat{C}')Y(\Omega, \mathbf{C}', P, P^k)}{P^k(\hat{C} \cup \hat{C}')}.$$

This postulate expresses the linearity property of the answer depth function, similarly to the analogous postulate for question difficulty (see [21]). We expect that it will be modified (or dropped) when more general models of information exchange are considered.

Just it was done for questions in [21], we say that the subset D of Ω is *homogeneous* iff the conditional depth function depends only on measures of partition $\mathbf{C}$ whenever $\hat{C} \subseteq D$, i.e. $Y(D, \mathbf{C}, P_D, P^k_D) = f(P_D(\mathbf{C}), P^k_D(\mathbf{C}))$. In particular, any atom (minimal set) of $\mathcal{F}$ is homogeneous. The next postulate concerns answers to free-response questions located inside homogeneous regions of the parameter space Ω.

Postulate A6 (*Homogeneous free-response sequentiality*). Let $D \subseteq \Omega$ be a homogeneous subset of the parameter space and let C be a free-response question such that $C \subseteq D$. Then

$$Y(\Omega, C, P, P^k) = Y(\Omega, D, P, P^k) + Y(D, C, P_D, P_D^k).$$

This postulate states that, whether a free-response question located inside a homogeneous region of the parameter space is answered in stages or right away, the overall effort required of the answerer (to achieve certain fixed accuracy) is the same. For example, let the question be *"What species does this animal belong to?"*. Instead of answering this question right away the source could answer a question about the order first, then suborder, then superfamily, family and only then about the actual species. In general, it is clear that the effort required to answer the original question right away could be more (i.e. if the animal is exotic) or less (i.e. if the animal is common like a domestic cat) than that required to answer the same question in stages to the same accuracy. Postulate A6 sates that the effort would be the same if all questions involved are located inside a homogeneous region. We expect that this postulate would be retained (perhaps in a modified form) when more general information exchange models are considered.

We can now state the main result about the possible shape of answer conditional depth function $Y(\Omega, \mathbf{C}, P, P^k)$. It is formulated as a theorem.

Theorem 2 *Let Postulates A1 through A6 hold. Then the conditional answer depth function* $Y(\Omega, \mathbf{C}, P, P^k)$ *has the following form*

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

where $u(C_j) = \frac{\int_{C_j} u(\omega) dP^k(\omega)}{P^k(C_j)}$ *and the integrable function* $u\colon \Omega \to \mathbb{R}$ *is the same that is used in characterizing the question difficulty function* $G(\cdot)$.

Proof: The proof is similar to that of main theorem in [21]. We can assume without loss of generality that there exists a complete partition $\mathbf{D} = \{D_1, \ldots, D_N\}$ of Ω such that every subset in $\mathbf{D}$ is homogeneous.

Let D be a homogeneous subset of Ω and let $C' \subset C \subset D$ be two subsets of D. Then, by Postulate A6,

$$Y(\Omega, C, P, P^k) = Y(\Omega, D, P, P^k) + Y(D, C, P_D, P_D^k), \tag{9}$$

and

$$Y(D, C', P_D, P^k_D) = Y(D, C, P_D, P^k_D) + Y(C, C', P_C, P^k_C). \tag{10}$$

Since D is homogeneous it follows from (10) that

$$f(P_D(C'), P^k_D(C')) = f(P_D(C), P^k_D(C)) + f(P_D(C')/P_D(C), P^k_D(C')/P^k_D(C)).$$

Then standard arguments using Postulates A1 and A2 (see [24] for details) lead to the conclusion that the function $f(\cdot)$ has the form

$$f(p, q) = c \log \frac{q}{p},$$

where c is a positive constant. Going back to the function Y we obtain

$$Y(D, C, P_D, P^k_D) = u'(D) \log \frac{P^k_D(C)}{P_D(C)}, \tag{11}$$

where $u'(D) > 0$ is a constant that can possibly depend on the particular homogeneous subset D.

Substituting (11) into (9) we arrive at

$$\begin{aligned} Y(\Omega, C, P, P^k) - Y(\Omega, D, P, P^k) = Y(D, C, P_D, P^k_D) &= u'(D) \log \frac{P^k_D(C)}{P_D(C)} \\ &= u'(D) \log \frac{P^k(C)}{P(C)} - u'(D) \log \frac{P^k(D)}{P(D)}, \end{aligned}$$

from which it follows (using continuity of Y and the fact that the subset $C \subset D$ is arbitrary) that

$$Y(\Omega, C, P, P^k) = u'(D) \log \frac{P^k(C)}{P(C)} + v'(D),$$

for any $C \subset D$ whenever D is homogeneous. Here $v'(D)$ is another constant that can possibly depend on the homogeneous subset D. We can now use Postulate A3 to conclude that $u'(D) = u(D)$ for all homogeneous sets D and that $v(D') \equiv 0$. This leads to the following expression for the conditional depth function of a free-response answer lying inside a homogenous subset:

$$Y(\Omega, C, P, P^k) = u(D) \log \frac{P^k(C)}{P(C)}. \tag{12}$$

Now let $\mathbf{D} = \{D_1, \ldots, D_N\}$ be a (complete) partition of Ω such that every subset in $\mathbf{D}$ is homogeneous. Let $C \subset \Omega$ be a free-response question. Using Postulate A5, we can write

$$Y(\Omega, \mathbf{D}_C, P, P^k) = \frac{\sum_{j=1}^N u(D_j) P^k(C \cap D_j) \log \frac{P^k(C \cap D_j)}{P(C \cap D_j)}}{P^k(C)}, \tag{13}$$

and

$$Y(C, \mathbf{D}_C, P_C, P^k_C) = \sum_{j=1}^{N} u(D_j) \frac{P^k(C \cap D_j)}{P^k(C)} \log \frac{P^k(C \cap D_j)/P^k(C)}{P(C \cap D_j)/P(C)}. \tag{14}$$

An application of Postulate A4 now yields

$$\begin{aligned} Y(\Omega, C, P, P^k) &= Y(\Omega, \mathbf{D}_C, P, P^k) - Y(C, \mathbf{D}_C, P_C, P^k_C) \\ &= \sum_{j=1}^{N} u(D_j) \frac{P^k(C \cap D_j)}{P^k(C)} \log \frac{P^k(C)}{P(C)} = u(C) \log \frac{P^k(C)}{P(C)}, \end{aligned}$$

where

$$u(C) \equiv \sum_{j=1}^{N} \frac{P^k(C \cap D_j) u(D_j)}{P^k(C)} = \frac{\int_C u(\omega) dP^k(\omega)}{P^k(C)}. \tag{15}$$

Here, the function u: $\Omega \to \mathbb{R}$ is defined as

$$u(\omega) = \sum_{j=1}^{N} u(D_j) I_{D_j}(\omega),$$

and therefore is the same exact function that was used to describe the question difficulty function G.

Finally, let $\mathbf{C} = \{C_1, \ldots, C_r\}$ be an arbitrary question on Ω. An application of Postulate A5 yields

$$Y(\Omega, \mathbf{C}, P, P^k) = \frac{\sum_{j=1}^{r} u(C_j) P^k(C_j) \log \frac{P^k(C_j)}{P(C_j)}}{\sum_{j=1}^{r} P^k(C_j)}, \tag{16}$$

where $u(C_j)$ is given by (15). □

Having found the expression for conditional depth function we can now use it to obtain the unconditional (expected) answer depth $Y(\Omega, \mathbf{C}, P, V(\mathbf{C}))$. We formulate the result as a corollary.

Corollary 1 *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 $u(C_j)$ is as defined in Theorem 2.

In the following, we will often use the notation $\Pr(V(\mathbf{C}) = s_k) \equiv v_k$ for the sake of brevity.

6 Relationship Between Difficulty and Depth

Theorems 1 and 2 (together with Corollary 1) establish the overall form that question difficulty and answer depth, respectively, can take. The conditional depth function $Y(\Omega, \mathbf{C}, P, P^k)$ depends, besides the original measure P, on the updated measure $P^k \equiv P^{V(\mathbf{C})=s_k}$.

6.1 Multiple-choice Questions

For multiple-choice (complete) questions, it makes sense to assume that the original measure P is a "valid" one in the sense that it does not change on average upon reception of the answer message $V(\mathbf{C})$. More formally speaking, for any question $\mathbf{C} = \{C_1, \ldots, C_r\}$,

$$P = \sum_{k=1}^{m} \Pr(V(\mathbf{C}) = s_k) P^k, \;\; j = 1, \ldots, r. \tag{17}$$

The expression (17) can be thought of as a condition of consistency of the answer message $V(\mathbf{C})$ with the original measure P and can be used for determining probabilities $v_k \equiv \Pr(V(\mathbf{C}) = s_k)$ of various values of the answer message $V(\mathbf{C})$.

Note that, taking into account the form (6) of the updated measures P^k the consistency condition (17) can be written as

$$\sum_{j=1}^{m} v_k p_{kj} = P(C_j), \;\; j = 1, \ldots, r. \tag{18}$$

Let us assume the consistency condition (17) holds and consider the answer depth function given by Corollary 1. Since for a multiple-choice question $\sum_{j=1}^{r} P^k(C_j) = 1$, we can write

$$\begin{aligned} Y(\Omega, \mathbf{C}, P, V(\mathbf{C})) &= \sum_{k=1}^{m} v_k \sum_{j=1}^{r} u(C_j) P^k(C_j) \log \frac{P^k(C_j)}{P(C_j)} \\ &= \sum_{k=1}^{m} v_k \sum_{j=1}^{r} u(C_j) P^k(C_j) \log P^k(C_j) - \sum_{k=1}^{m} v_k \sum_{j=1}^{r} u(C_j) P^k(C_j) \log P(C_j) \\ &\overset{(a)}{=} \sum_{k=1}^{m} v_k \sum_{j=1}^{r} u(C_j) P^k(C_j) \log P^k(C_j) + G(\Omega, \mathbf{C}, P) \overset{(b)}{\leq} G(\Omega, \mathbf{C}, P), \end{aligned}$$

where (a) follows from (17) and Theorem 1, and (b) follows from the inequality $\log P^k(C_j) \leq 0$. It is also clear that the inequality (b) becomes an equality if and only if, for every value s_k of the answer message, either $P^k(C_j) = 0$ or $\log P^k(C_j) = 0$ for every value of the index j. For the latter to be true it is necessary and sufficient that, for all values of k,

$$P^k(C_j) = \delta_{f(k),j}, \tag{19}$$

where f: $\{1, 2, \ldots, m\} \to \{1, 2, \ldots, r\}$ is a map from the set of possible values of index k to that of index j. Substituting (19) into (17) we obtain

$$P(C_j) = \sum_{k=1}^{m} v_k \delta_{f(k),j} = \sum_{k:f(k)=j} v_k. \tag{20}$$

It is easy to see that without loss of generality one can define an equivalent message $V'(\mathbf{C})$ such that $V'(\mathbf{C}) = s_j$ whenever $V(\mathbf{C}) = s_k$ such that $f(k) = j$. Then (20) becomes simply

$$P(C_j) = \Pr(V'(\mathbf{C}) = s_j). \tag{21}$$

A *perfect* answer to a multiple-choice question is defined $\mathbf{C} = \{C_1, \ldots, C_r\}$ as the message $V(\mathbf{C}) = \{s_1, \ldots, s_r\}$ such that $P^k(C_j) = \delta_{k,j}$, and, as a consequence, $\Pr(V(\mathbf{C}) = s_j) = P(C_j)$. Then we can state the result obtained above as a lemma.

Lemma 1 *Let $\mathbf{C}$ be a multiple-choice question and assume the condition (17) for any answer $V(\mathbf{C})$ holds. Then $Y(\Omega, \mathbf{C}, P, V(\mathbf{C})) \leq G(\Omega, \mathbf{C}, P)$ with the inequality being tight if and only if the answer $V(\mathbf{C})$ is perfect (up to trivial equivalences).*

6.2 Free-response Questions

Let $C \subset \Omega$ be a free-response question. We can write the depth function for a corresponding answer $V(C)$ as follows.

$$\begin{aligned} Y(\Omega, C, P, V(C)) &= \sum_{k=1}^{m} \Pr(V(C) = s_k) u(C) P^k(C) \log \frac{P^k(C)}{P(C)} \\ &= u(C) \sum_{k=1}^{m} \Pr(V(C) = s_k) \log P^k(C) - u(C) \log P(C) \sum_{k=1}^{m} \Pr(V(C) = s_k) \\ &= u(C) \sum_{k=1}^{m} \Pr(V(C) = s_k) \log P^k(C) + G(\Omega, C, P) \overset{(a)}{\leq} G(\Omega, C, P), \end{aligned}$$

where (a) follows from that the inequality $\log P^k(C) \leq 0$. It is straightforward to see that for the inequality (a) to become an equality it is necessary and sufficient that $P^k(C) = 1$ for all values k

of the answer message. Clearly, in that case, we can define an equivalent message $V'(C)$ that takes a single value s so that $P^s(C) = 1$.

A *perfect* answer $V(C)$ to a free-response question C is defined to be a message taking a single value s such that $P^s(C) = 1$.

We can again state the result obtained above as a lemma.

Lemma 2 *Let C be a free-response question and $V(C)$ an answer to it. Then $Y(\Omega, C, P, V(C)) \leq G(\Omega, C, P)$ with the inequality being tight if and only if the answer $V(C)$ is perfect (up to trivial equivalences).*

6.3 Mixed Questions

Finally, let $\mathbf{C} = \{C_1, \ldots, C_r\}$ where $\hat{C} = \cup_{i=j}^r C_j \subset \Omega$ be a mixed question. We define a *perfect* answer $V(\mathbf{C})$ to a mixed question as a message taking values in the set $\{s_1, \ldots, s_r\}$ such that $P^j(C_j) = 1$ for $j = 1, \ldots, r$.

For any answer to a mixed question we demand that the following *consistency with the original knowledge* condition holds.

$$\sum_{k=1}^{m} \Pr(V(\mathbf{C}) = s_k) P^k(C_j) = \gamma P(C_j), \tag{22}$$

where

$$\gamma = \frac{P^s(\hat{C})}{P(\hat{C})} = \frac{\sum_{j=1}^r P^k(C_j)}{\sum_{j=1}^r P(C_j)} \tag{23}$$

for all values of k characterizes the free-response component of $V(\mathbf{C})$. Then it is straightforward to prove a result analogous to that of Lemmas 1 and 2.

Lemma 3 *If $\mathbf{C}$ is a mixed question and $V(\mathbf{C})$ is an answer to it such that condition (22) holds. Then $Y(\Omega, \mathbf{C}, P, V(\mathbf{C})) \leq G(\Omega, \mathbf{C}, P)$ with the inequality becoming tight if and only if the answer $V(\mathbf{C})$ is perfect.*

Proof: We can write the depth function for $V(\mathbf{C})$ as follows.

$$
\begin{aligned}
Y(\Omega, \mathbf{C}, P, V(\mathbf{C})) &= \sum_{k=1}^{m} v_k \frac{\sum_{j=1}^{r} u(C_j) P^k(C_j) \log \frac{P^k(C_j)}{P(C_j)}}{\sum_{j=1}^{r} P^k(C_j)} \\
&\overset{(a)}{=} \frac{\sum_{k=1}^{m} \sum_{j=1}^{r} v_k u(C_j) P^k(C_j) \log \frac{P^k(C_j)}{P(C_j)}}{\gamma P(\hat{C})} \\
&= \frac{1}{\gamma P(\hat{C})} \sum_{k=1}^{m} \sum_{j=1}^{r} v_k u(C_j) P^k(C_j) \log P^k(C_j) \\
&- \frac{1}{\gamma P(\hat{C})} \sum_{k=1}^{m} \sum_{j=1}^{r} v_k u(C_j) P^k(C_j) \log P(C_j) \\
&\overset{(b)}{=} \frac{1}{\gamma P(\hat{C})} \sum_{k=1}^{m} \sum_{j=1}^{r} v_k u(C_j) P^k(C_j) \log P^k(C_j) - \frac{1}{P(\hat{C})} \sum_{j=1}^{r} u(C_j) P(C_j) \log P(C_j) \\
&= \frac{1}{\gamma P(\hat{C})} \sum_{k=1}^{m} \sum_{j=1}^{r} v_k u(C_j) P^k(C_j) \log P^k(C_j) + G(\Omega, \mathbf{C}, P) \overset{(c)}{\leq} G(\Omega, \mathbf{C}, P),
\end{aligned}
$$

where (a) follows from (23), (b) follows from (22) and (c) follows from the inequality $\log P^k(C_j) \leq 0$. Using the same arguments as those employed for the proof of Lemma 1 we arrive at the statement of this lemma. □

7 Quasi-perfect Answers to Complete Questions

Let the question $\mathbf{C} = \{C_1, \ldots, C_r\}$ be complete (multiple-choice) and let $V(\mathbf{C})$ be an answer to $\mathbf{C}$. If $V(\mathbf{C})$ is perfect, its depth $Y(\Omega, \mathbf{C}, P, V(\mathbf{C}))$ is equal to the difficulty $G(\Omega, \mathbf{C}, P)$ of $\mathbf{C}$ as Lemma 1 states. Here we would like to consider some simple classes of imperfect answers. To make the form of an imperfect answer more specific let us assume such an answer to resemble a perfect one in that the number of possible values it can take is equal to r and each message s_k, $k = 1, \ldots, r$ expresses a degree of preference towards the subset C_k. Let e_k be the error probability associated with s_k, i.e $e_k = P^k(\bar{C}_k)$, where $\bar{C}_k = \Omega \backslash C_k$. Let us also make the additional assumption that the error associated with s_k is "proportionally distributed" between the sets C_j $j \neq k$, i.e. $P^k(C_j) = \frac{e_k P(C_j)}{P(\bar{C}_k)} = \frac{e_k P(C_j)}{1 - P(C_k)}$. Obviously, both of these assumptions can be stated in the following way.

$$
P^k = (1 - e_k) P_{C_k} + \sum_{j \neq k} \frac{e_k P(C_j)}{1 - P(C_k)} P_{C_j},
$$

implying that the coefficients p_{kj} in (6) have the form

$$p_{kj} = \left(1 - \frac{e_k}{1 - P(C_k)}\right)\delta_{k,j} + \frac{e_k P(C_j)}{1 - P(C_k)} \tag{24}$$

To further simplify the analysis and provide more concise description of errors associated with imperfect answers, we make a further assumption: that the error probability e_k constitutes the same fraction of $P(\bar{C}_k)$ for all values of k, i.e. $e_k = \alpha(1 - P(C_k))$, $k = 1, \ldots, r$, where $0 \leq \alpha \leq 1$. Under this assumption, the error associated with the answer $V(\mathbf{C})$ that we will denote by $V_\alpha(\mathbf{C})$ is fully described by a single parameter α. The coefficients p_{kj} in (24) become

$$p_{kj} = (1 - \alpha)\delta_{k,j} + \alpha P(C_j), \tag{25}$$

and the updated measure P^k becomes simply

$$P^k = \alpha P + (1 - \alpha) P_{C_k}. \tag{26}$$

We see that for $\alpha = 0$ the measure P^k turns into the conditional measure P_{C_k} making the answer perfect, and for $\alpha = 1$ each measure P^k becomes the original measure P thus rendering the answer $V_\alpha(\mathbf{C})$ empty, i.e. possessing vanishing depth.

Substituting (26) into the general expression for the answer depth and using the fact that in this case $v_k = P(C_k)$, $k = 1, \ldots, r$, we can obtain

$$\begin{aligned} Y(\Omega, \mathbf{C}, P, V_\alpha(\mathbf{C})) = \sum_{k=1}^{r} u(C_k) P(C_k)(1 - \alpha + \alpha P(C_k)) \log \frac{1 - \alpha + \alpha P(C_k)}{P(C_k)} \\ + \alpha \log \alpha \sum_{k=1}^{r} u(C_k) P(C_k)(1 - P(C_k)), \end{aligned} \tag{27}$$

It is easy to see that the expression (27) becomes $G(\Omega, \mathbf{C}, P)$ for $\alpha = 0$ and vanishes for $\alpha = 1$.

In the following we will call answers characterized by updated measures of the form (26) and depth functions given by (27) the *quasi-perfect* answers. Their advantage is that they allow to smoothly interpolate between perfect and empty answers using just a single parameter α taking values on the interval $[0, 1]$.

Substituting (25) into the consistency condition (18) it is easy to see that for quasi-perfect answers

$$v_j = P(C_j), \tag{28}$$

for $j = 1, \ldots, r$, regardless of the value of error probability α.

8 Relationships Between Questions

Given two complete questions $\mathbf{C}'$ and $\mathbf{C}''$ the *pseudo-energy overlap* $J(\Omega, (\mathbf{C}'; \mathbf{C}''), P)$ was defined in [21] as

$$J(\Omega, (\mathbf{C}'; \mathbf{C}''), P) = G(\Omega, \mathbf{C}', P) + G(\Omega, \mathbf{C}'', P) - G(\Omega, \mathbf{C}' \cap \mathbf{C}'', P), \tag{29}$$

which can easily be seen to have the following form

$$J(\Omega, (\mathbf{C}'; \mathbf{C}''), P) = \sum_{i=1}^{r'} \sum_{j=1}^{r''} u(C_i' \cap C_j'') P(C_i' \cap C_j'') \log \frac{P(C_i' \cap C_j'')}{P(C_i')P(C_j'')}. \tag{30}$$

It was also shown that the pseudo-energy overlap can be interpreted as the reduction of difficulty of question $\mathbf{C}''$ due to the knowledge of a perfect answer $V^*(\mathbf{C}')$ to question $\mathbf{C}'$.

$$G(\Omega, \mathbf{C}'', V^*(\mathbf{C}')) = G(\Omega, \mathbf{C}'', P) - J(\Omega, (\mathbf{C}'; \mathbf{C}''), P), \tag{31}$$

where the *conditional difficulty* $G(\Omega, \mathbf{C}'', V(\mathbf{C}'))$ is defined (for any answer $V^*(\mathbf{C}')$ to question $\mathbf{C}'$) as

$$G(\Omega, \mathbf{C}'', V^*(\mathbf{C}')) = \sum_{k=1}^{m'} \Pr(V^*(\mathbf{C}') = s_k) G(\Omega, \mathbf{C}'', P'^k). \tag{32}$$

It would be interesting to find out how the relation (31) generalizes for the case of an arbitrary answer to question $\mathbf{C}'$. Clearly, since a reception of value s_k' of $V(C')$ updates the measure P to P'^k, the difficulty of $\mathbf{C}''$ given $V(C') = s_k'$ is equal to

$$\begin{aligned} G(\Omega, \mathbf{C}'', P'^k) &= -\sum_{j=1}^{r''} u(C_j'') P'^k(C_j'') \log P'^k(C_j'') \\ &= -\sum_{j=1}^{r''} \sum_{l=1}^{r'} u(C_l' \cap C_j'') P'^k(C_l' \cap C_j'') \log P'^k(C_j''), \end{aligned}$$

and therefore the overall (expected) difficulty $G(\Omega, \mathbf{C}'', V(\mathbf{C}'))$ of question $\mathbf{C}''$ given an answer

$V(\mathbf{C}')$ to $\mathbf{C}'$ can be written – denoting $\Pr(V(\mathbf{C}') = s'_k)$ by v'_k – as

$$G(\Omega, \mathbf{C}'', V(\mathbf{C}')) \equiv \sum_{k=1}^{m'} v'_k G(\Omega, \mathbf{C}'', P'^k) = -\sum_{k=1}^{m'} v'_k \sum_{j=1}^{r''} \sum_{l=1}^{r'} u(C'_l \cap C''_j) P'^k(C'_l \cap C''_j) \log P'^k(C''_j)$$

$$= \sum_{k=1}^{m'} v'_k \left(\sum_{j=1}^{r''} \sum_{l=1}^{r'} u(C'_l \cap C''_j) P'^k(C'_l \cap C''_j) \log P'^k(C''_j) \right.$$

$$\left. + \sum_{j=1}^{r''} \sum_{l=1}^{r'} u(C'_l \cap C''_j) P'^k(C'_l \cap C''_j) \log P(C''_j) - \sum_{j=1}^{r''} \sum_{l=1}^{r'} u(C'_l \cap C''_j) P'^k(C'_l \cap C''_j) \log P(C''_j) \right)$$

$$= -\sum_{k=1}^{m'} v'_k \sum_{j=1}^{r''} \sum_{l=1}^{r'} u(C'_l \cap C''_j) P'^k(C'_l \cap C''_j) \log \frac{P'^k(C''_j)}{P(C''_j)}$$

$$- \sum_{j=1}^{r''} \sum_{l=1}^{r'} u(C'_l \cap C''_j) P'^k(C'_l \cap C''_j) \log P(C''_j)$$

$$= G(\Omega, \mathbf{C}'', P) - \sum_{k=1}^{m'} v'_k \sum_{j=1}^{r''} \sum_{l=1}^{r'} u(C'_l \cap C''_j) P'^k(C'_l \cap C''_j) \log \frac{P'^k(C''_j)}{P(C''_j)}. \tag{33}$$

We see from (33) that the conditional difficulty of $\mathbf{C}''$ can be represented as a difference of the standard (unconditional) difficulty and another expression that can be appropriately denoted $Y(\Omega, \mathbf{C}'', P, V(\mathbf{C}'))$ and called the *relative depth* of the answer $V(\mathbf{C}')$ with respect to question $\mathbf{C}''$:

$$G(\Omega, \mathbf{C}'', V(\mathbf{C}')) = G(\Omega, \mathbf{C}'', P) - Y(\Omega, \mathbf{C}'', P, V(\mathbf{C}')), \tag{34}$$

where the relative depth $Y(\Omega, \mathbf{C}'', P, V(\mathbf{C}'))$ is given by

$$Y(\Omega, \mathbf{C}'', P, V(\mathbf{C}')) = \sum_{k=1}^{m'} v'_k \sum_{j=1}^{r''} \sum_{l=1}^{r'} u(C'_l \cap C''_j) P'^k(C'_l \cap C''_j) \log \frac{P'^k(C''_j)}{P(C''_j)}. \tag{35}$$

Using the expression (6) for the updated measures P'^k we find that

$$P'^k(C'_l \cap C''_j) = p_{kl} \frac{P(C'_l \cap C''_j)}{P(C'_l)} \tag{36}$$

and

$$P'^k(C''_j) = \sum_{l=1}^{r'} p_{kl} \frac{P(C'_l \cap C''_j)}{P(C'_l)}, \tag{37}$$

and, substituting (36) and (37) into (35) we obtain for the relative depth:

$$Y(\Omega, \mathbf{C}'', P, V(\mathbf{C}')) = \sum_{k=1}^{m'} v'_k \sum_{l=1}^{r'} \sum_{j=1}^{r''} u(C'_l \cap C''_j) p'_{kl} \cdot \frac{P(C'_l \cap C''_j)}{P(C'_l)} \log \sum_{i=1}^{r'} p'_{ki} \cdot \frac{P(C'_i \cap C''_j)}{P(C'_i) \cdot P(C''_j)}. \tag{38}$$

We can summarize the result just obtained as a lemma.

Lemma 4 *Let $\mathbf{C}'$ and $\mathbf{C}''$ be two arbitrary complete questions on Ω and let $V(\mathbf{C}')$ be an answer to $\mathbf{C}'$. Then the conditional difficulty of $\mathbf{C}''$ given the answer $V(\mathbf{C}')$ can be found as*

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

where the relative depth of $V(\mathbf{C}')$ is given by the expression (38).

Suppose now that $V^*(\mathbf{C}')$ is a perfect answer to $\mathbf{C}'$ which implies that $m' = r'$ and $p'_{kl} = \delta_{k,l}$. Substituting this into (38) and performing the sum over k while making use of the answer consistency condition (18) we obtain

$$Y(\Omega, \mathbf{C}'', P, V^*(\mathbf{C}')) = \sum_{l=1}^{r'} \sum_{j=1}^{r''} u(C'_l \cap C''_j) P(C'_l \cap C''_j) \log \frac{P(C'_l \cap C''_j)}{P(C'_l)P(C''_j)}, \tag{39}$$

which coincides with the expression (30) for the pseudo-energy overlap between questions $\mathbf{C}'$ and $\mathbf{C}''$. We thus recover the result (31) obtained in [21].

Let now $V_\alpha(\mathbf{C}')$ be a quasi-perfect answer to question $\mathbf{C}'$ characterized by error probability α. Substituting expressions (25) and (28) into (38) we obtain, after some straightforward algebra

$$\begin{aligned} Y(\Omega, \mathbf{C}'', P, V_\alpha(\mathbf{C}')) &= (1-\alpha) \sum_{l=1}^{r'} \sum_{j=1}^{r''} u(C'_l \cap C''_j) P(C'_l \cap C''_j) \log \left[(1-\alpha) \frac{P(C'_l \cap C''_j)}{P(C'_l)P(C''_j)} + \alpha \right] \\ &+ \alpha \sum_{l=1}^{r'} \sum_{j=1}^{r''} u(C'_l \cap C''_j) P(C'_l \cap C''_j) \sum_{k=1}^{r'} P(C'_k) \log \left[(1-\alpha) \frac{P(C'_k \cap C''_j)}{P(C'_k)P(C''_j)} + \alpha \right] \end{aligned} \tag{40}$$

It is easy to see that for $\alpha = 0$ (40) reduces to (39) which is the overlap between questions $\mathbf{C}'$ and $\mathbf{C}''$, and for $\mathbf{C}''$ coinciding with $\mathbf{C}'$ the relative depth (40) becomes the depth $Y(\Omega, \mathbf{C}', P, V_\alpha(\mathbf{C}'))$ (given by expression (27)) of quasi-perfect answer to $\mathbf{C}'$ characterized by the same value of error probability α. To see that it is sufficient to set $C''_j = C'_j$ (and hence $P(C'_l \cap C''_j) = \delta_{l,j} P(C'_l)$) in (40) and make use of the (obvious) identity $\sum_{k \neq j} P(C'_k) = 1 - P(C'_j)$.

9 Examples

Let us revisit the example with a finite parameter space from [21]. The parameter space Ω consists of 8 elements, corresponding to green, yellow and red apples (denoted GA, YA and RA, respectively), green, yellow and red pears (denoted GPr, YPr and RPr), and yellow and red peaches (denoted YPc and RPc). The elements are equiprobable so that $P(\cdot) = \frac{1}{8}$ for all $\omega \in \Omega$. The function

$u(\omega)$ describes the relative difficulty of respective free-response questions. The observation that the green (cold) color is easier to distinguish from both the yellow and red (warm) colors is reflected in $u(GA) = u(GPr) = 1$. On the other hand, an apple and a pear are also easy to distinguish from each other because of a different overall shape. (Recall that there is no green peach that could possibly be confused with a green apple.) The observation that yellow and red pears can possibly be confused with each other but not with anything else because of either their warm color (compared to green pears) or their distinct shape (compared to red or yellow apples or peaches) is reflected in $u(YPr) = u(RPr) = 1.5$. Finally, $u(YA) = u(RA) = u(YPc) = u(RPc) = 2$ as these four combinations appear to be the hardest to distinguish from each other as they all possess a warm color and round shape. Normalizing the values of $u(\cdot)$ so that $\int_\Omega u(\omega)dP(\omega) = 1$ we obtain $u(GA) = u(GPr) = \frac{8}{13}$, $u(YPr) = u(RPr) = \frac{12}{13}$ and $u(YA) = u(RA) = u(YPc) = u(RPc) = \frac{16}{13}$.

Consider, as in [21], the question *"Is the fruit green or not?"*. Let $C_g = \{GA, GPr\} \subset \Omega$ be the subset consisting of all green fruit (apples and pears) and let $\overline{C}_g = \Omega \setminus C_g$ be the subset containing fruit of all other colors (red and yellow). The partition is $\mathbf{C}_g = \{C_g, \overline{C}_g\}$. The values $u(\cdot)$ for the sets in this partition are $u(C_g) = \frac{8}{13}$ and $u(\overline{C}_g) = \frac{1}{3} \cdot \frac{12}{13} + \frac{2}{3} \cdot \frac{16}{13} = \frac{44}{39}$. The measures are $P(C_g) = \frac{1}{4}$ and $P(\overline{C}_g) = \frac{3}{4}$. The second similar question is *"Is the fruit a peach or not?"*. The corresponding partition is $\mathbf{C}_{Pc} = \{C_{Pc}, \overline{C}_{Pc}\}$ where $C_{Pc} = \{YPc, RPc\}$ and $\overline{C}_{Pc} = \Omega \setminus C_{Pc}$. The values of function $u(\cdot)$ on these subsets are $u(C_{Pc}) = \frac{16}{13}$ and $u(\overline{C}_{Pc}) = \frac{1}{3} \cdot \frac{8}{13} + \frac{1}{3} \cdot \frac{12}{13} + \frac{1}{3} \cdot \frac{16}{13} = \frac{12}{13}$. The measures are $P(C_{Pc}) = \frac{1}{4}$ and $P(\overline{C}_{Pc}) = \frac{3}{4}$. Let $V_\alpha(\mathbf{C}_g)$ and $V_\alpha(\mathbf{C}_{Pc})$ be the corresponding quasi-perfect answers. The depth functions of these answers can be computed using (27) as (see Fig. 3 for an illustration)

$$Y(\Omega, \mathbf{C}_g, P, V_\alpha(\mathbf{C}_g)) = \frac{2}{13}\left(1 - \frac{3}{4}\alpha\right)\log(4 - 3\alpha) + \frac{11}{13}\left(1 - \frac{1}{4}\alpha\right)\log\frac{4-\alpha}{3} + \frac{15}{52}\alpha\log\alpha,$$

and

$$Y(\Omega, \mathbf{C}_{Pc}, P, V_\alpha(\mathbf{C}_{Pc})) = \frac{4}{13}\left(1 - \frac{3}{4}\alpha\right)\log(4 - 3\alpha) + \frac{9}{13}\left(1 - \frac{1}{4}\alpha\right)\log\frac{4-\alpha}{3} + \frac{21}{52}\alpha\log\alpha.$$

Consider the question *"What color is the given fruit?"* on one hand and *"What type is the given fruit?"* on the other. The former question can be represented as the partition $\mathbf{C}_c = \{C_g, C_y, C_r\}$ where $C_g = \{GA, GPr\}$, $C_y = \{YA, YPr, YPc\}$ and $C_r = \{RA, RPr, RPc\}$; the latter question can be identified with the partition $\mathbf{C}_t = \{C_A, C_{Pr}, C_{Pc}\}$ where $C_A = \{GA, YA, RA\}$, $C_{Pr} = \{GPr, YPr, RPr\}$ and $C_{Pc} = \{YPc, RPc\}$. The values of $u(\cdot)$ on these subsets are $u(C_g) = \frac{8}{13}$, $u(C_y) = \frac{1}{3} \cdot \frac{12}{13} + \frac{2}{3} \cdot \frac{16}{13} = \frac{44}{39}$, $u(C_g) = u(C_y) = \frac{44}{39}$; $u(C_A) = \frac{1}{3} \cdot \frac{8}{13} + \frac{2}{3} \cdot \frac{16}{13} = \frac{40}{39}$, $u(C_{Pr}) = \frac{1}{3} \cdot \frac{8}{13} + \frac{2}{3} \cdot \frac{12}{13} =$

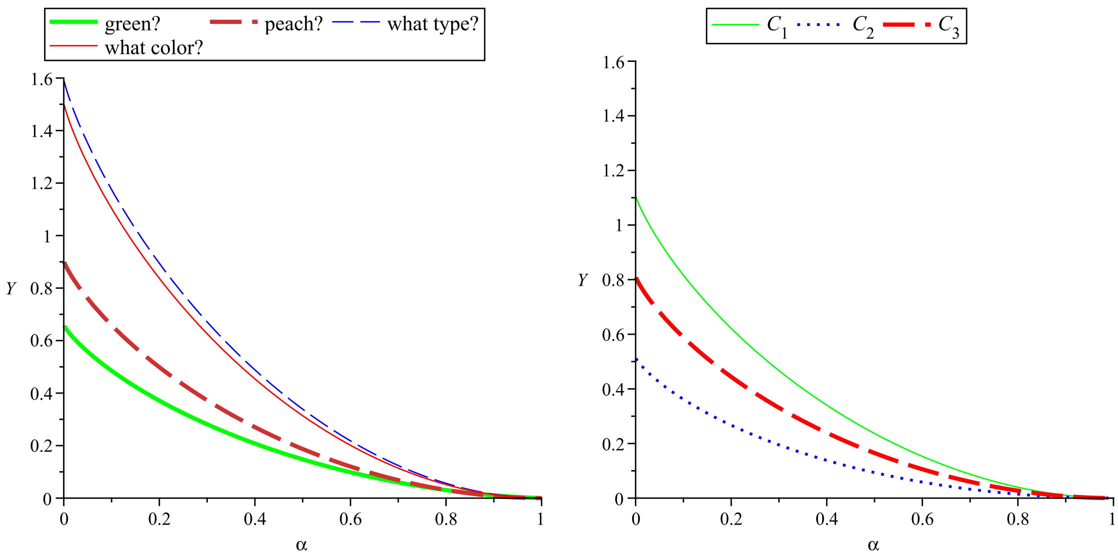

Figure 3: Answer depth as function of α for quasi-perfect answers to questions on the finite parameter space (left) and infinite parameter space (right).

$\frac{32}{39}$, $u(C_{Pc}) = \frac{16}{13}$. The measures are $P(C_g) = \frac{1}{4}$, $P(C_y) = \frac{3}{8}$, $P(C_r) = \frac{3}{8}$; $P(C_A) = P(C_{Pr}) = \frac{3}{8}$, $P(C_{Pc}) = \frac{1}{4}$. Let $V_\alpha(\mathbf{C}_c)$ and $V_\alpha(\mathbf{C}_t)$ be quasi-perfect answers to questions $\mathbf{C}_c$ and $\mathbf{C}_t$. The depth of these answers can be found using the expression (27). The results are (see Fig. 3)

$$Y(\Omega, \mathbf{C}_c, P, V_\alpha(\mathbf{C}_c)) = \frac{2}{13}\left(1 - \frac{3}{4}\alpha\right)\log(4-3\alpha) + \frac{11}{13}\left(1 - \frac{5}{8}\alpha\right)\log\frac{8-5\alpha}{3} + \frac{67}{104}\alpha\log\alpha,$$

and

$$Y(\Omega, \mathbf{C}_t, P, V_\alpha(\mathbf{C}_t)) = \frac{4}{13}\left(1 - \frac{3}{4}\alpha\right)\log(4-3\alpha) + \frac{9}{13}\left(1 - \frac{5}{8}\alpha\right)\log\frac{8-5\alpha}{3} + \frac{69}{104}\alpha\log\alpha.$$

Let us consider the second example from [21]. The parameter space is $\Omega = [0,1]^2 \subset \mathbb{R}^2$. Let the pseudo-temperature function be $u(\omega) = \frac{3}{2}(\omega_1^2 + \omega_2^2)$ (so that the hard questions are located towards the upper-right corner of Ω). Consider the following three subsets of Ω: $C_1 = \{\omega : \omega_1 \in [\frac{1}{2}, 1], \omega_2 \in [\frac{1}{2}, 1]\}$, $C_2 = \{\omega : \omega_1 \in [0, \frac{1}{2}], \omega_2 \in [0, \frac{1}{2}]\}$, $C_3 = \{\omega : \omega_1 \in [0, \frac{1}{2}], \omega_2 \in [\frac{1}{2}, 1]\}$ and let $\mathbf{C}_i = \{C_i, \overline{C}_i\}$ for $i = 1, 2, 3$ be three complete questions on Ω. Let $V(\mathbf{C}_i)$ be a quasi-perfect answer to question $\mathbf{C}_i$, $i = 1, 2, 3$ characterized by error probability α. We can use the expression (27) to obtain the depth of these answers (see Fig. 3 for an illustration).

$$Y(\Omega, \mathbf{C}_1, P, V(\mathbf{C}_1)) = \frac{7}{16}\left(1 - \frac{3}{4}\alpha\right)\log(4-3\alpha) + \frac{9}{16}\left(1 - \frac{1}{4}\alpha\right)\log\frac{4-\alpha}{3} + \frac{15}{32}\alpha\log\alpha,$$

$$Y(\Omega, \mathbf{C}_2, P, V(\mathbf{C}_2)) = \frac{1}{16}\left(1 - \frac{3}{4}\alpha\right)\log(4-3\alpha) + \frac{15}{16}\left(1 - \frac{1}{4}\alpha\right)\log\frac{4-\alpha}{3} + \frac{9}{32}\alpha\log\alpha,$$

and

$$Y(\Omega, \mathbf{C}_3, P, V(\mathbf{C}_3)) = \frac{1}{4}\left(1 - \frac{3}{4}\alpha\right)\log(4-3\alpha) + \frac{3}{4}\left(1 - \frac{1}{4}\alpha_1\right)\log\frac{4-\alpha}{3} + \frac{3}{8}\alpha\log\alpha.$$

In all these examples, we see that, as expected, the depths of answers to more difficult questions is higher for the same accuracy (value of error probability α). In other words, it takes more effort on the part of the information source to answer a more difficult question with the same accuracy. Equivalently, the same amount of effort (measured by pseudo-energy) yields a lower accuracy answer to a more difficult question. We can see from Fig. 3 that, for instance, a quasi-perfect answer of depth equal to 0.4 to the question *"Is the fruit green?"* has an error probability of around 0.18, but an equally deep (i.e. of the same depth) answer to the more difficult question *"Is the fruit a peach or not?"* has a larger error probability of around 0.24.

Let us turn to relative depth of answers. Consider the above example again. The relative depth $Y(\Omega, \mathbf{C}'', P, V_\alpha(\mathbf{C}'))$ of a quasi-perfect answer $V_\alpha(\mathbf{C}')$ with respect to question $\mathbf{C}''$ can be readily computed using the expression (40). We obtain, for questions $\mathbf{C}_1$ and $\mathbf{C}_2$,

$$\begin{aligned} Y(\Omega, \mathbf{C}_2, P, V_\alpha(\mathbf{C}_1)) &= \left(\frac{1}{2} - \frac{7}{32}\alpha\right)\log\left(\frac{4}{3}(1-\alpha)+\alpha\right) \\ &+ \left(\frac{1}{2} + \frac{13}{64}\alpha\right)\log\left(\frac{8}{9}(1-\alpha)+\alpha\right) + \frac{1}{64}\alpha\log\alpha, \end{aligned}$$

$$\begin{aligned} Y(\Omega, \mathbf{C}_1, P, V_\alpha(\mathbf{C}_2)) &= \left(\frac{1}{2} - \frac{1}{32}\alpha\right)\log\left(\frac{4}{3}(1-\alpha)+\alpha\right) \\ &+ \left(\frac{1}{2} - \frac{5}{64}\alpha\right)\log\left(\frac{8}{9}(1-\alpha)+\alpha\right) + \frac{7}{64}\alpha\log\alpha. \end{aligned}$$

Likewise, for questions $\mathbf{C}_1$ and $\mathbf{C}_3$, we have

$$\begin{aligned} Y(\Omega, \mathbf{C}_3, P, V_\alpha(\mathbf{C}_1)) &= \left(\frac{11}{16} - \frac{5}{16}\alpha\right)\log\left(\frac{4}{3}(1-\alpha)+\alpha\right) \\ &+ \left(\frac{5}{16} + \frac{1}{4}\alpha\right)\log\left(\frac{8}{9}(1-\alpha)+\alpha\right) + \frac{1}{16}\alpha\log\alpha, \end{aligned}$$

and

$$\begin{aligned} Y(\Omega, \mathbf{C}_1, P, V_\alpha(\mathbf{C}_3)) &= \left(\frac{11}{16} - \frac{7}{32}\alpha\right)\log\left(\frac{4}{3}(1-\alpha)+\alpha\right) \\ &+ \left(\frac{5}{16} + \frac{7}{64}\alpha\right)\log\left(\frac{8}{9}(1-\alpha)+\alpha\right) + \frac{7}{64}\alpha\log\alpha. \end{aligned}$$

Finally, for questions $\mathbf{C}_2$ and $\mathbf{C}_3$, expression (40) yields

$$Y(\Omega, \mathbf{C}_3, P, V_\alpha(\mathbf{C}_2)) = \left(\frac{5}{16} + \frac{1}{16}\alpha\right) \log\left(\frac{4}{3}(1-\alpha) + \alpha\right) \\ + \left(\frac{11}{16} - \frac{1}{8}\alpha\right) \log\left(\frac{8}{9}(1-\alpha) + \alpha\right) + \frac{1}{16}\alpha \log \alpha,$$

and

$$(\Omega, \mathbf{C}_2, P, V_\alpha(\mathbf{C}_3)) = \left(\frac{5}{16} - \frac{1}{32}\alpha\right) \log\left(\frac{4}{3}(1-\alpha) + \alpha\right) \\ + \left(\frac{11}{16} + \frac{1}{64}\alpha\right) \log\left(\frac{8}{9}(1-\alpha) + \alpha\right) + \frac{1}{64}\alpha \log \alpha.$$

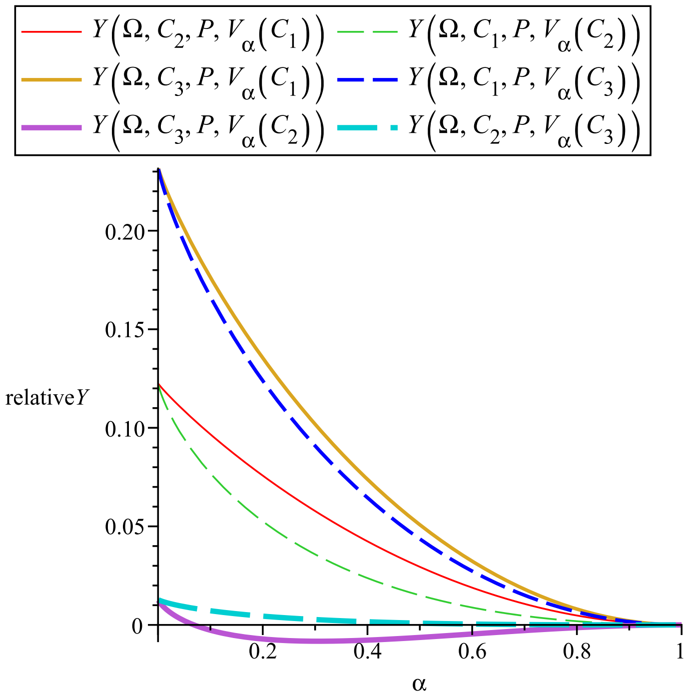

Figure 4: Relative depth of quasi-perfect answers as function of α.

These relative depth curves are shown in Fig. 4. We can see, in particular, that the relative depth $Y(\Omega, \mathbf{C}'', P, V_\alpha(\mathbf{C}'))$ is not in general symmetric in the two questions unless $\alpha = 0$ or $\alpha = 1$. In the former case the relative depth reduces to the overlap $J(\Omega, (\mathbf{C}'; \mathbf{C}''), P)$ which is symmetric and in the latter case the relative depth simply vanishes. Further, it can be seen from Fig. 4 that the relative depth can in fact be negative meaning that it is possible that the knowledge of an (imperfect) answer to a question may make another question more difficult. It would be interesting to establish general conditions under which relative depth is nonnegative. Another useful observation is that if for a pair of questions $\mathbf{C}'$ and $\mathbf{C}''$, question $\mathbf{C}'$ is the more difficult one of the two then it appears that the inequality $Y(\Omega, \mathbf{C}'', P, V_\alpha(\mathbf{C}')) > Y(\Omega, \mathbf{C}', P, V_\alpha(\mathbf{C}''))$ holds for $0 < \alpha < 1$ implying that a quasi-perfect answer to a more difficult question results in a higher reduction of difficulty of the other question. It would be of interest to see if this property holds in the general case or exceptions are possible.

10 Conclusion

This article continues the development of a quantitative framework for optimization of the process of additional information acquisition in problems of decision making under uncertainty that was initiated in [21]. The proposed framework contains three main ingredients: the information source itself, questions that the source can receive (from a decision maker) and answers that the source can give in response to questions. Questions were addressed in [21] where the general definition of a question was given and the concept of question difficulty function as a quantitative measure of the degree of difficulty of a given question to a given source was introduced. It was noted that the question difficulty can – using parallels with thermodynamics – be thought of as the amount of pseudo-energy associated with the question. The overall form of the question difficulty function was established in [21] by demanding that it satisfy a number of requirements (called postulates) that express some reasonable properties the difficulty function can be expected to satisfy. The system of postulates proposed and used in [21] can – using the parallels with thermodynamics mentioned above – be termed “ideal gas model”, specifically in that it expresses the linearity and isotropy properties of the difficulty function. It was shown in [21] that the most general overall shape of the difficulty function satisfying the postulates of the ideal gas model is described by a scalar function on the problem parameter space that can be interpreted as temperature-like quantity and that was termed pseudo-temperature.

The subject of the present article is answers that the information source can give in response to questions. In particular, any answer to a question can be assigned the amount of pseudo-energy that measures the answer depth, i.e. the amount of “work” the source has to do in order to answer the question to the given accuracy. Clearly, the higher the desired accuracy is the more “work” the source would have to do and the higher the answer depth is. Also, if properly defined, it makes sense that the answer depth has to be bounded by the question difficulty from above reaching that bound if and only if the answer is fully correct. In this article, the overall form of the answer depth function was established in the way similar to how it was done for the question difficulty in [21]. Namely, reasonable postulates were formulated that the answer depth function had to satisfy. Just like in [21], the proposed system of postulates expressed the linearity and isotropy properties of the answer depth function. One can say therefore that the latter is obtained within the "ideal gas model" that was already used in [21]. It turns out that the resulting depth function is described, besides appropriate probability measures, by a scalar function on the problem parameter space that has to be the same pseudo-temperature function describing the corresponding question difficulty.

In addition to answer depth, the relative depth of an answer to one question with respect to another question was defined that can be used to determine how an answer to one question reduces the difficulty of a different question. It is expected that the relative depth will become especially useful when the optimal additional information acquisition process with multiple information sources is studied in later publications.

*Bibliography

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

[2] 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.

[3] 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.

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

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

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

[7] Faddeev, D. K. 1956. On the concept of entropy of a finite probabilistic scheme. *Uspekhi Mat. Nauk* **11**(1(67)) 227–231.

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

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

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

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

[12] Havrda, J. H., F. Charvat. 1967. Quantification methods of classification processes: Concepts of structural α entropy. *Kybernetika* **3** 30–35.

[13] Jaroszewicz, S., D. A. Simovici. 1999. On axiomatization of conditional entropy of functions between finite sets. *Proc. 29th ISMVL, Freiburg, Germany*. 24–31.

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

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

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

[17] 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.

[18] 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.

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

[20] 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.

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

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

[23] 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.

[24] Rényi, A. 1961. On measures of entropy and information. *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I*. Univ. California Press, Berkeley, Calif., 547–561.

[25] Simovici, D. A., S. Jaroszewicz. 2002. An axiomatization of partition entropy. *IEEE Trans. Inf. Theory* **48**(7) 2138–2142.

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

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