Information Acquisition for Decision Making: Question Difficulty

David Grace
Lehigh University

Eugene Perevalov
Lehigh University

Report: 12T-001

Information Acquisition for Decision Making: Question Difficulty

Draft

Abstract

The problem of optimal decision making in environments characterized with both uncertainty and presence of information sources is considered in a general setting. This motivates searching for quantitative measures of question difficulty that would allow for maximizing the effect of additional information the information sources are capable of supplying. In this article, the concept of question difficulty for questions identified with partitions of problem parameter space is introduced and the overall form of question difficulty function is derived that satisfies a particular system of reasonable postulates. It is found that the resulting difficulty function depends on a single scalar function on the parameter space that can be interpreted – using parallels with classical thermodynamics – as temperature-like quantity, with the question difficulty being similar to thermal energy. Quantitative relationships between different questions are also explored.

1 Introduction

In decision making under uncertainty – depending on the particular features of the problem at hand – different methods are used. Problems for which the main difficulty lies in a large number of possible solutions and a complex structure of the feasible region are usually considered to belong to the domain of optimization (stochastic [4], robust [2, 1] or, more recently, risk-averse [12, 38, 3]). The state of knowledge (typically expressed by a probability distribution on the space of problem parameters) about the unknown problem parameters is usually assumed to be fixed. If the number of possible solutions is relatively small and the main difficulty lies in the process of updating the initial information (probability distribution), decision theoretic methods are appropriate. In Markov decision processes and stochastic optimal control, respectively, additional assumptions (such as Markovian or Gaussian property) are made which allows one to obtain solutions with special properties making it possible to handle the time aspect of the problem efficiently.

In many practical decision making problems, sources of additional information are often available but the knowledge they possess is typically not structured well and thus cannot be used directly in, for instance, stochastic optimization formulations. In addition, such information is often highly imprecise, noisy and is characterized by various degrees of relevance to the problem the decision maker is facing. That further complicates the task of extracting the additional information and using it optimally from the point of view of solution quality improvement. A good example of such a situation is the product portfolio selection problem some electronics manufacturing companies are regularly facing. The number of products they are typically choosing from is measured in dozens making the number of possible combinations enormous. In addition, multiple constraints on feasible combinations exist complicating the feasible region structure. The initial reliable information about possible future gains associated with various products is typically very scarce. On the other hand, it is almost universally believed that many people in the organization possess valuable additional information that fails to be utilized by the decision makers due to reasons just mentioned.

The present article initiates development of a unified theoretical framework for optimal information acquisition in general purpose decision making problems including those with large and complex feasible regions. The proposed framework begins with a consideration of a decision making problem on one hand and an information source on the other. The source is assumed to be capable of answering questions that are related to the input data of the decision making problem. Generally speaking, any source has finite capability that manifests itself in answering easy questions with higher accuracy than difficult ones. Difficulty of various questions is source-specific: what is easy for one source can be difficult for another and vice versa. On the other hand, different questions have different degree of relevance for the given decision making problem: having one question answered accurately may increase the solution quality significantly while possessing an equally accurate answer to another – equally detailed – question may have a much smaller effect on the solution quality. This naturally leads to an important question the decision maker appears to be facing: how the information source should be optimally "aligned" with the given problem, or, more precisely, what question the decision maker should ask the information source so that the respective answer would have the largest positive effect on the solution quality for the given problem. More generally, if several information sources are available, the decision maker would want to know what question(s) and – possibly – in what order should be asked the sources so that the combined effect of the respective answers on the solution quality can be maximized. In other words, here the overall problem is that of optimal "alignment" of a system of information sources to the given decision making problem. What can make that latter problem more difficult is that optimal question(s) to be asked a given source might in general depend on the number and properties ("expertise") of other available sources.

Development of such a framework clearly has to include several components:

- Model(s) of an information source that, in particular, allows for an evaluation of relative difficulty of various questions (the source's "expertise");
- Definition of a question and a measure of its difficulty;
- Definition of an answer to a given question and a measure of its accuracy;
- Description of a relationship of source's answers to various questions and the solution quality for the given decision making problem;
- Development of efficient solution methods for the optimal "alignment" problem mentioned earlier.

Since a model of an information source has to describe the accuracy of the source's answers to various questions it appears natural to begin with definitions of questions and answers and their respective quantitative characteristics – that can be called *difficulty* and *depth*, respectively. The present article is devoted to the former of these tasks: the development of a quantitative theory of questions.

1.1 Related Work

This and the following papers can be described as an attempt to bring information theory to bear on optimization and decision making. As is well-known, the field of Information Theory that grew out of Shannon's pioneering work on communication theory has since had a profound impact on a number of disciplines in natural sciences and engineering. One of fundamental advances brought by Information Theory is the concept of entropy and mutual information that provide a natural and consistent measure of the amount of information associated with general probability distributions. Besides a revolution in communications which started from the demonstration that error-free transmission over imperfect channel was possible and gave rise to the modern coding theory, the list of successful applications of these concepts includes (but is not limited to) a simple derivation of statistical physics laws [24, 25], new algorithms in computer vision [43], new methods of analysis in climatology [33, 42], physiology [27] and neurophysiology [7]. The latter were based on the concept of transfer entropy [39] which can interpreted as conditional mutual information defined for time series and that can be used to measure information transfer between different parts of complex systems. The relatively new field of Generalized Information Theory (see e.g. [28]) is concerned with problems of characterizing uncertainty in frameworks that are more general than classical probability such as Dempster-Shafer theory [40]. There it was shown, for example, that the minimal uncertainty measure satisfying consistency requirements such as general subadditivity and additivity for combining uncertainty for independent subsystems is obtained by maximizing Shannon entropy over all classical probability distributions consistent with the given belief specification [31, 21]. This paper generalizes the concept of entropy by introducing the concept of *pseudo-energy* the purpose of which is to quantitatively characterize the information source "expertise" and that can be thought of as an amount of "work" associated with a given question.

The approach proposed here and in companion papers [34, 35] can be thought of as a development of a general theory of inquiry going back to the work of Cox [10, 11] that received more attention recently. One of the main results of these later developments is the *calculus of inquiry* [29] that, in particular, constructs a distributive lattice of questions dual to the Boolean lattice of logical assertions. The definition of questions adapted in this paper, in fact, corresponds to the particular subclass of questions defined in [29] – the partition questions. Our work here and in [34, 35] goes somewhat further by introducing the concept of pseudo-energy as a measure of source specific difficulty of various questions to the given information source. It is possible to say that we develop a quantitative theory of *knowledge* as opposed to the theory of information.

In this paper, we use an axiomatic approach to determine the overall form of the question difficulty function. The latter can be thought of as a logical development of the entropy concept of information theory. The axiomatic approach was first used, besides Shannon himself, in [14] to derive the most general form of the entropy function. Later, [37] used a different set of axioms to find the one-parameter family of functions (later called Rényi entropies) that included standard (Shannon) entropy as a special case. The concept of structural entropy was introduced in [22] and used for classification purposes. Also known as Havrda-Charvat entropy, it was more recently obtained by axiomatic means in [41] where axiomatization of partition entropy was discussed on rather general grounds (see also [23] for closely related work). It was shown in [41] that Shannon entropy, Havrda-Charvat entropy and Gini index all obtain as particular cases of general partition entropy that satisfies a system of reasonable axioms.

The main motivation for this and following companion papers [34] and [35] is, as was already mentioned, to develop methodology for optimal use of (additional) information in decision making under uncertainty. This idea obviously not entirely new and it has been studied and used, for instance, in the area of statistical decision making. Applications to innovation adoption [32, 26], fashion decisions [16] and vaccine composition decisions for flu immunization [30] can be mentioned in this regard. It's interesting to observe that the amount of information in these applications is typically measured simply as the number of relevant observations which can either be costless or costly, depending on the model. Some authors [15, 13] introduced various models (e.g. effective information model) for accounting for the actual, or effective, amount of information contained in the received observations. The common theme of this line of work is to try to find an optimal trade-off between the amount of additional information obtained and the suitably measured degree of achieving the original goal. Thus, for instance, in [30], waiting longer allows the decision makers to obtain more precise forecast of which flu virus strains are going to be predominant but leaves less time for actual vaccine production. The main difference of the approach initiated by this paper is in that it allows to optimize not only the quantity of the acquired information but also its content and that it explicitly accounts for properties of information sources.

Explicit consideration of information sources that lies at the core of the proposed methodology is similar in spirit to analyzing and using information provided by human experts. In fact, in many practically relevant applications the role of multi-purpose information sources used in the proposed approach will likely be played by experts. In existing research 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 probability assessment from multiple experts [18, 19, 8, 9] and, in particular, optimal combining of expert opinions, including experts with incoherent and missing outputs [36]. Especially closely related to the approach initiated in this paper are the investigations on using and combining information of experts that partition the event differently [5] and on rules of updating probabilities based on outcomes of partially similar events [6]. 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 [17]. In the approach developed in this and consecutive papers, the emphasis is on *optimizing* on the particular type of information (i.e. partition) for the given expert(s) and the given decision making problem.

1.2 Outline

The rest of the article is organized as follows. In the next section, details about the main motivation for the proposed framework are given. Section 3 contains necessary technical preliminaries. In section 4, the main components of the proposed framework are described. Section 5 is devoted to a discussion of the question difficulty function – the main topic of the present article. In particular, the main theorem establishing the overall shape of the question difficulty function that is required to satisfy certain reasonable postulates is proved. In section 6, relationships between different questions are explored. Section 7 contains simple numerical examples illustrating the results obtained earlier in the article. Finally, a conclusion summarizing the main results is given in section 8.

2 Motivation: Decision Making Under Uncertainty

In decision making under uncertainty, the goal is to choose the best decision given the available information, according to a suitable criterion. One of the most widely used criteria is that of optimizing the *expected* objective function given the probability distribution that describes the available information. The problem so formulated can be formally written as

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

Here $X \subset \mathcal{D}$ is the set of all *feasible* solutions, i.e. the set satisfying all (deterministic) constraints that are present in the problem formulation, where $\mathcal{D}$ is the space to which all solutions belong (e.g. a suitable Euclidean space). Ω has the meaning of a space of possible values of input data parameters that are not known with certainty. It is often referred to as a parameter space. P is a fixed initial probability measure (with a suitable sigma-algebra assumed) on Ω that describes the initial state of the uncertainty and that can in principle be modified by querying information sources. The function $f\colon \Omega \times \mathcal{D} \to \overline{\mathbb{R}}$ is assumed to be integrable on Ω for each $x \in X$. For example, in the context of stochastic optimization, X is the set of feasible first-stage solutions and $f(\omega, x)$ is the best possible objective value for the first stage decision x in case when the random outcome ω is observed.

We are interested, given the problem (1) and an information source capable of providing answers to our questions, in obtaining the best possible solution to problem (1), suitably modified by the source's answer(s). To make this desideratum a bit more specific, let $L(P)$ be the *expected loss* corresponding to measure P defined 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 ω.

Let $\mathfrak{Q}$ be the set of all possible (suitably defined) questions that can be directed towards the source of information, and let $A(Q)$ be its answer to a particular question $Q \in \mathfrak{Q}$. Further, let P_a be the measure on Ω conditional on reception of a particular value a of the answer A. One can think of P_a as the measure updated by the value a, from the original measure P. Then the expected loss following question Q and answer $A = A(Q)$ can be found as

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

where the sum is over all possible values a of the answer A.

Our goal then can be stated as that of finding, for the given problem (1) and a given information source, the question(s) $Q \in \mathfrak{Q}$ that would make the corresponding expected loss (2) as small as possible:

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

Informally speaking, the problem is about finding the question(s) that is "aligned" optimally with both the information source's "strengths" and the particular decision making problem. Changing the purely "optimization" component of the problem (the function $f(\omega, x)$ and the set X) while keeping the "information" component (the space Ω and the measure P) the same will in general change the optimal question(s) Q for the same information source. Thus the main goal can also be described as that of finding an optimal alignment between the optimization and information components of the problem (where the information source itself is included in the latter).

3 Preliminaries

In the following we denote by Ω the base space consisting of all possible outcomes of potential interest to the decision maker. We will often refer to it, as mentioned earlier, as parameter space. Ω can be finite or infinite, such as a closed subset of a Euclidean space $\mathbb{R}^s$. We denote by $\mathcal{F}$ a sigma-algebra on Ω. Let P be a fixed probability measure on $(\Omega, \mathcal{F})$. We will usually refer to it – and other measures – as a measure on Ω, omitting an explicit specification of $\mathcal{F}$ unless needed.

Let $C \in \mathcal{F}$ be a (measurable) subset of Ω. We denote by P_C the conditional measure on Ω 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 $\tilde{\mathbf{C}}$. Given measure P on Ω, we call partition $\mathbf{C}_f(P)$ the *finest* partition of Ω associated with measure P if $P(C) > 0$ for all $C \in \mathbf{C}_f(P)$ and there exists at least one set of zero measure in any refinement of $\mathbf{C}_f(P)$. In case Ω is a closed subset of a Euclidean space and $\mathcal{F}$ is a Borel algebra, it is easy to see that finest partitions do not exist if measure P has a continuous support or has a component with continuous support. It is also clear that if the measure P has discrete support there exist many partitions of Ω that are finest for P.

Let $\mathbf{C}' = \{C'_1, \ldots, C'_r\}$ and $\mathbf{C}'' = \{C''_1, \ldots, C''_s\}$ be two partitions of Ω. Then 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). Obviously, some of the sets constituting partition $\mathbf{C}' \cap \mathbf{C}''$ may be empty. Clearly, partition $\mathbf{C}' \cap \mathbf{C}''$ is a refinement of both $\mathbf{C}'$ and $\mathbf{C}''$.

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

Besides standard partitions of Ω, we will also need *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 will use the notation $\hat{C} \equiv \cup_{i=1}^r C_i$. Clearly, partition $\mathbf{C}$

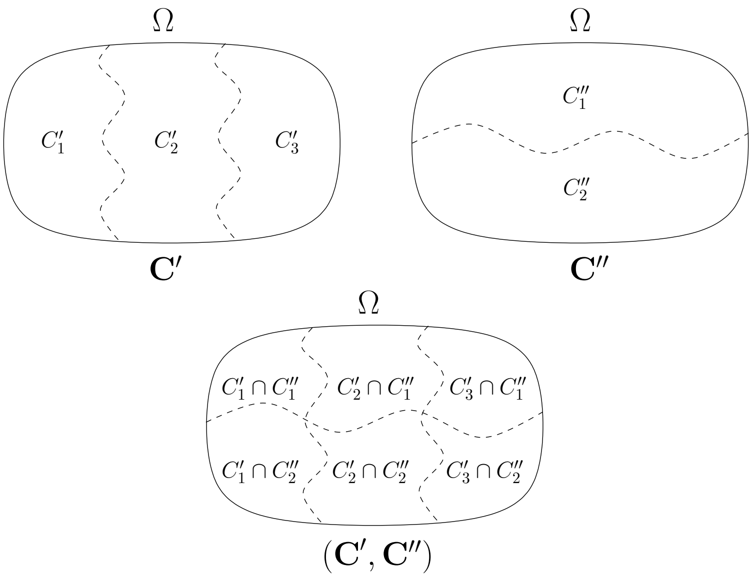

Figure 1: Two partitions of Ω and the corresponding joint partition.

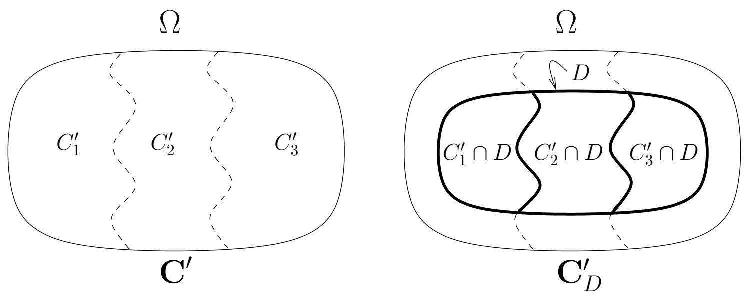

Figure 2: Partition $\mathbf{C}_D$ of set $D \subset \Omega$ induced by a partition $\mathbf{C}$ of Ω.

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 that are completely disjoint, i.e. such that $\hat{C}' \cap \hat{C}'' = \emptyset$. Then the partition $\mathbf{C} = \mathbf{C}' \cup \mathbf{C}''$ is defined as partition consisting of all subsets in the constituent partitions: $\mathbf{C} = \{C'_1, \ldots, C'_r, C''_1, \ldots, C''_s\}$. Clearly, partition $\mathbf{C}' \cup \mathbf{C}''$ may be complete or incomplete (it would be complete if and only if $\hat{C}' \cup \hat{C}'' = \Omega$). In case $\hat{C}' \cap \hat{C}'' \neq \emptyset$, the partition $\mathbf{C}' \cup \mathbf{C}''$ is not defined.

For an arbitrary complete partition $\mathbf{C} = \{C_1, \ldots, C_r\}$, it is straightforward to show that 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 Ingredients

As was mentioned earlier, the main components of the information exchange framework developed here are information sources, decision maker's questions and corresponding source's answers. Below, we discuss them in turn with some emphasis on questions which are the main subject of the present article.

4.1 Information Source

Assume that a source of additional information is available that is capable of answering specific questions concerning input data for problem (1). This implies that the source's answers are capable of modifying the initial measure P on Ω. The overall idea that we would like to formalize can be summarized as a set of – loosely formulated at this point – reasonable assumptions.

- 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. The result is that it answers questions well (with low error probabilities) if the question difficulty does not exceed its capacity and the quality of its answers progressively degrades as the difficulty exceeds the source's capacity.

4.2 Questions

A question is a request for new information on top of what is already known. The latter is represented by the measure P on the parameter space Ω and is assumed to be common knowledge. Since, in the context discussed, any information is represented by some measure on Ω – with a measure concentrated on a single element of Ω corresponding to a state of full knowledge – a question can be associated to a specific request for an updated measure on Ω. Therefore we identify a question with a (possibly incomplete) *partition* of Ω.

Definition: A question is a partition $\mathbf{C} = \{C_1, C_2, \ldots, C_r\}$ where C_j, $j = 1, \ldots, r$ are subsets of Ω such that $C_i \cap C_j = \emptyset$ for $i \neq j$ and $\cup_{j=1}^{r} C_j \subseteq \Omega$.

Note that we allow for incomplete partitions for which $\cup_{j=1}^{r} C_j$ is a proper subset of Ω. For any partition $\mathbf{C}$, we denote the union of all subsets in $\mathbf{C}$ by $\hat{C}$: $\hat{C} \equiv \cup_{j=1}^{r} C_j$. Thus for any complete partition $\mathbf{C}$, $\hat{C} = \Omega$.

In everyday terms, a complete partition can be interpreted as a *multiple-choice* question (e.g. *"Is this apple red, green or yellow?"*. An incomplete partition consisting of a single subset can be associated with a *free-response* question, e.g. *"What color is this apple?"* Incomplete partitions consisting of several subsets can be interpreted as combinations of these two kinds – as *mixed* questions, e.g. *"What fruit is it and is it red, green or yellow?"*. In the given more narrow context – when the parameter space Ω and measure P on it are precisely known to the information source – the interpretation of an incomplete partition as a free-response (or mixed) question is not quite correct since if the source is presented with a description of a subset C of Ω the question becomes implicitly multiple-choice: *"Is the random outcome ω in C or not?"*. In order to accurately model real free-response questions (as they are usually understood), more complicated models are likely needed. From the more narrow point of view adapted in this article, incomplete questions are best thought of as an auxiliary construction that helps in determining the difficulty of complete questions that can be unambiguously defined and interpreted.

In the following, we will use the terms "partition" and "question" interchangeably. We will also use terms "multiple-choice question", "free-response question" and "mixed question" to mean complete partition, incomplete partition consisting of a single set and incomplete partitions consisting of more than one set, respectively.

Given a question $\mathbf{C}$, we are interested in quantifying its degree of difficulty, i.e. finding, for the given parameter space Ω and measure P on Ω, a function G: $\mathbf{C} \to \mathbb{R}$ that assigns larger values to more difficult questions. We use the notation $G(\Omega, \mathbf{C}, P)$ to emphasize the dependence of the question difficulty on Ω and P.

The particular shape of the function $G(\Omega, \mathbf{C}, P)$ could conceivably range in a fairly broad domain and would have to be approximated and estimated using experimental data such as observed performance of a source on various questions. On the other hand, due to the very fact of possibly wide range of shapes of the question difficulty function it makes sense to try to limit that range somewhat by imposing reasonable restrictions on the properties of the difficulty function. Such imposed restrictions can naturally be termed *postulates*. Then the validity of such postulates can be tested by observing a source's performance (such as empirical error probabilities) on various question of this type.

4.3 Perfect Answers

While a detailed discussion of answers will be given in the companion paper [34], here we introduce a concept of a *perfect answer* to a question $\mathbf{C}$ as an answer that provides an exhaustive reply to $\mathbf{C}$. Specifically, we have the following definition.

Definition: Given a question $\mathbf{C} = \{C_1, \dots, C_r\}$, the perfect answer $V^*(\mathbf{C})$ is a message that takes one of the values in the set $\{s_1, \dots, s_r\}$ such that the measure $P^j \equiv P^{V^*(\mathbf{C})=s_j}$ updated by a reception of value s_j of $V^*(\mathbf{C})$ is equal to the conditional measure P_{C_j}.

Informally speaking, a perfect answer to $\mathbf{C}$ completely resolves the uncertainty associated with the partition $\mathbf{C}$, i.e. places a random outcome ω in one of the subsets in $\mathbf{C}$ with certainty but otherwise does no more (since the resulting measure on the subset C_j is the conditional measure P_{C_j}). One can say that a perfect answer is the most basic type of an answer to a given question. It is convenient to think of a question difficulty $G(\Omega, \mathbf{C}, P)$ as an amount of *pseudo-energy* (the term 'motivated by certain parallels with thermodynamics) contained in $\mathbf{C}$. Then it is natural to require that the *depth* of a perfect answer $V^*(\mathbf{C})$ be equal to the difficulty of $\mathbf{C}$. In other words, the amount of pseudo-energy contained in a perfect answer to $\mathbf{C}$ is equal to that in $\mathbf{C}$.

If question $\mathbf{C}$ is complete and $V^*(\mathbf{C})$ is the corresponding perfect answer it is reasonable to assume that $V^*(\mathbf{C})$ does not change the original measure P on average, or, in other words, that the original measure P is a "valid" one that only gets refined by the answer to $\mathbf{C}$. Formally speaking, this assumption means that

$$\sum_{j=1}^{r} \Pr(V^*(\mathbf{C}) = s_j) P_{C_j} = P, \tag{6}$$

from which it follows – by evoking (5) – that $\Pr(V^*(\mathbf{C}) = s_j) = P(C_j)$. We will call (6) the *consistency* condition for answer $V^*(\mathbf{C})$.

For incomplete question, the condition (6) has to be modified to read

$$\sum_{j=1}^{r} \Pr(V^*(\mathbf{C}) = s_j) P_{C_j} = P_{\hat{C}}, \tag{7}$$

from which it follows that $\Pr(V^*(\mathbf{C}) = s_j) = \frac{P(C_j)}{P(\hat{C})}$.

5 Question Difficulty Function

Our goal in this section is to derive a general form of the question difficulty function $G(\Omega, \mathbf{C}, P)$ and – along the way – establish the set of parameters it can depend upon. In many fields of scientific inquiry, when faced with a new phenomenon, *linear* models are often explored first both because of their simplicity and because of their role as elementary building blocks for more complicated models. We will attempt to do same in our situation. Besides linearity, we will – somewhat implicitly – assume that the parameter space is *isotropic*, i.e. the pseudo-energy amount does not depend on the orientation of subsets in $\mathbf{C}$ in the parameter space. Later on, both of these basic assumption – linearity and isotropy – can be relaxed and more general models can be obtained.

As has been mentioned earlier, in the model adapted here, incomplete (free-response and mixed) questions are to be understood as auxiliary constructions, while complete (multiple-choice) questions have a clear meaning. For a free-response question $C \subset \Omega$, the difficulty function $G(\Omega, C, P)$ can be interpreted as *conditional* difficulty of any complete question $\mathbf{C}$ containing the subset C given that the random outcome ω is in C. For example, if the subset C_1 represents apple, C_2 – pear and C_3 - peach so that $C_1 \cup C_2 \cup C_3 = \Omega$, then $G(\Omega, C_1, P)$ can be interpreted as the difficulty of the question *"Is it an apple, a pear, or a peach?"*, or, equivalently *"What kind of fruit is it?"* (since the source knows that the possible types are apple, pear and peach), provided that an apple is shown to the source.

One reasonable and almost obvious requirement that can be imposed on the question difficulty function $G(\Omega, \mathbf{C}, P)$ is that of *certainty*, i.e. the difficulty of a question should vanish if there is no new knowledge to acquire given the original state of it. Formally speaking, $G(\Omega, \mathbf{C}, P) = 0$ whenever $P(C_j) = 1$ for some value of the index j. One can say that in this case the question is already answered at the time of its formulation. These are questions of the kind *"Is this red apple red, green or yellow?"* for complete (multiple-choice) questions or *"What color is this red apple?"* for incomplete (free-response) questions. Thus we obtain

Postulate Q1 (*Certainty*). Suppose $\mathbf{C} = \{C_1, \ldots, C_r\}$ and $P(C_j) = 1$ for some value of j. Then $G(\Omega, \mathbf{C}, P) = 0$.

Note that Postulate Q1 implies neither linearity nor isotropy and should be included even if these two basic assumptions are relaxed.

The second postulate we propose is of the same universal variety. It simply requires that the question difficulty function be continuous in all its arguments (which are yet to be determined).

Postulate Q2 (*Continuity*). The function $G(\Omega, \mathbf{C}, P)$ is continuous in all its arguments.

Again, it seems to be reasonable to keep Postulate Q2 even if more general models are desired.

The next postulate states that, for questions that have both free-response and multiple-choice components, i.e. for questions that are incomplete but consist of several subsets, the difficulty is additive: the overall difficulty of the question is the sum of the difficulty of the free-response part and the difficulty of the multiple-choice part given the free-response part has been answered perfectly. Formally, we obtain the following.

Postulate Q3 (*Mixed question decomposition*). Let $\mathbf{C} = \{C_1, \ldots, C_r\}$ be such that $\hat{C} = \cup_{j=1}^r C_j \neq \Omega$. Then

$$G(\Omega, \mathbf{C}, P) = G(\Omega, \hat{C}, P) + G(\hat{C}, \mathbf{C}, P_{\hat{C}}).$$

This postulate describes the difficulty of questions of the sort *"What kind of fruit is it and is it red, green or yellow?"*. It states that the difficulty of the overall question is additive: it is equal to the sum of difficulties of two questions: *"What fruit is it?"* and *"Is this apple red, green or yellow?"* assuming the correct answer to the first question was *"Apple"*. This postulate may likely be changed or relaxed when more general models are considered.

The next postulate states the mean value property of incomplete questions: the difficulty of the question $\mathbf{C} \cup \mathbf{C}'$ obtained by taking the union of two incomplete non-overlapping partitions $\mathbf{C}$ and $\mathbf{C}'$ is equal to the arithmetic mean value of the difficulties of the constituents questions with respect to the original measure P.

Postulate Q4 (*Mean value*). Let $\mathbf{C}$ and $\mathbf{C}'$ be two incomplete questions such that $\hat{C} \cap \hat{C}' = \emptyset$. Then

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

This postulate can be interpreted as follows. Let $\mathbf{C}$ and $\mathbf{C}'$ each consist of a single subset: $\mathbf{C} = \{C\}$ and $\mathbf{C}' = \{C'\}$ for $C \subset \Omega$, $C' \subset \Omega$. Assume also that $C \cup C' = \Omega$, so that $\{C, C'\}$ is a complete question. Then the statement of Postulate Q4 would read

$$G(\Omega, \{C, C'\}, P) = P(C)G(\Omega, C, P) + P(C')G(\Omega, C', P), \tag{8}$$

which is consistent with the interpretation of the difficulty $G(\Omega, C, P)$ of a free-response question as difficulty of a multiple-choice question containing C as one of possible answers given that C is true (that is conditioned on $\omega \in C$). For instance, let C represent an apple and C' a pear and assume these are the only two possible types of fruit. Then expression (8) states that the difficulty of the question *"What kind of fruit is it?"* (which, given the structure of Ω and the measure P is equivalent to the question *"Is it an apple or a pear?"*) is equal to the difficulty of the same question in case an apple is shown times the probability that an apple can be shown plus the same expression for the pear. Thus $G(\Omega, \{C, C'\}, P)$ is the expected value of the multiple-choice question difficulty where the expectation is taken over possible correct answers. From this point of view, Postulate Q4 sounds rather natural and generic. However, the real meaning of Postulate Q4 is in that is states that the conditional difficulties are *independent of the number and measures of other options (subsets).* Postulate Q4 assigns the same conditional difficulty $G(\Omega, C, P)$ to the subset $C \subset \Omega$ *regardless of the complete partition it is a member of.* For instance, if $C \subset \Omega$ represents an apple then, in the case the source is shown an apple, the difficulty of the question *"Is it an apple or not?"* would be the same as that of *"What kind of a fruit is it?"* even if the number of possible choices (types of fruit) is large. It is easy to see that this, while not unreasonable, still is a rather strong assumption which may not be true for realistic information sources. Postulate Q4 can be thought of as an expression of linearity of the difficulty function and it can be fully expected that it will be relaxed or modified in more general models.

To state the next postulate we need to introduce a new concept. We say that the parameter space Ω is *homogeneous* if the question difficulty function depends only on its subset measures for any question $\mathbf{C}$ in Ω: $G(\Omega, \mathbf{C}, P) = f(P(\mathbf{C}))$ where $P(\mathbf{C})$ stands for the vector $(P(C_1), \ldots, P(C_r))$. More generally, we say that a subset $D \subseteq \Omega$ is *homogeneous* if $G(D, \mathbf{C}, P_D) = f(P_D(\mathbf{C}))$ as long as $\hat{C} \subseteq D$. In particular, any atom (minimal set) of the sigma-algebra $\mathcal{F}$ is homogeneous. Postulate 5 then states that a free-response question can be posed in stages without changing its overall difficulty as long as all the intermediate questions lie inside a homogeneous subset of the parameter space.

Postulate Q5 (*Homogeneous free-response sequentiality*). Let $D \subset \Omega$ be a homogeneous subset of the parameter space and let $\mathbf{C}$ be a question such that $\hat{C} \subset D$. Then

$$G(\Omega, \mathbf{C}, P) = G(\Omega, D, P) + G(D, \mathbf{C}, P_D).$$

To get a little more "feel" for this postulate think of a question asking to identify a certain animal species. The gradual approach to such a question would involve asking intermediate questions about the class the animal belongs to, order, suborder, superfamily, family, and, finally, the species itself. In case the original question is of "harder than average" variety it would be easier to answer the question in stages compared to answering it right away. On the other hand, if the original question is an easy one (easier than other similar questions) it can be easier to answer it without resorting to the intermediate "guiding" questions. A good example of the latter would be a question about a domestic cat that an average person would be able to answer easily and correctly whereas the "guiding" questions about class, order etc. would likely present some difficulty. Respectively, if all such questions are equally hard (for the same measure) then it would make sense to believe that the intermediate "guiding" questions would not change the difficulty of the original question just like the postulate states.

Finally, it certainly makes sense to require that if $D \subset \Omega$ is homogeneous and $C \subset D$ then $f(P_D(C)) = G(\Omega, C, P_D)$ should be a decreasing function of its argument $P_D(C)$. Indeed, a free-response question about something "rare" should be more difficult. We thus obtain Postulate 6.

Postulate Q6 (*Homogeneous free-response monotonicity*). Suppose $D \subset \Omega$ is homogeneous and $C \subset D$. Then $f(P_D(C)) = G(\Omega, C, P_D)$ is a decreasing function of its argument $P_D(C)$.

In order to get still more insight into the proposed set of postulates for the question difficulty function consider the following alternative postulate.

Postulate Q3' (*Multiple-choice sequentiality*). Let $\mathbf{C} = \{C_1, \ldots, C_r\}$ be a complete question and let $\tilde{\mathbf{C}}$ be its refinement. Then

$$G(\Omega, \tilde{\mathbf{C}}, P) = G(\Omega, \mathbf{C}, P) + \sum_{j=1}^{r} P(C_j) G(C_j, \tilde{\mathbf{C}}_{C_j}, P_{C_j}).$$

Postulate Q3' states that if a multiple-choice question is made more detailed the difficulty of the resulting question can be obtained as a sum of the difficulty of the original question and the average (with respect to the measure P) of difficulties of conditional detalizations. For instance if the original question was *"Is it an apple or a pear?"* and the detalization sounds like *"Is it an apple or a pear and is its color red, green or yellow?"* then Postulate Q3' says that the difficulty of the detailed question is equal to the difficulty of the original question plus the average of difficulties of questions *"Is this apple red, green or yellow?"* and the question *"Is this pear red, green or yellow?"*. This postulate may seem to be somewhat more reasonable and grounded in experience compared to, for instance, the *Mean value* postulate. It turns out though that Postulate Q3' is implied by Postulate Q3 and Postulate Q4 as the following lemma shows.

Lemma 1 *Suppose Postulate Q3 and Postulate Q4 hold. Then Postulate Q3' holds as well.*

Proof: Let $\tilde{\mathbf{C}}$ be a refinement of $\mathbf{C} = \{C_1, \ldots, C_r\}$. Then we can write

$$\begin{aligned} G(\Omega, \tilde{\mathbf{C}}, P) &\overset{(a)}{=} \sum_{j=1}^{r} P(C_j) G(\Omega, \tilde{\mathbf{C}}_{C_j}, P) \\ &\overset{(b)}{=} \sum_{j=1}^{r} P(C_j)(G(\Omega, C_j, P) + G(C_j, \tilde{\mathbf{C}}_{C_j}, P_{C_j})) \\ &= \sum_{j=1}^{r} P(C_j) G(\Omega, C_j, P) + \sum_{j=1}^{r} P(C_j) G(C_j, \tilde{\mathbf{C}}_{C_j}, P_{C_j}) \\ &\overset{(c)}{=} G(\Omega, \mathbf{C}, P) + \sum_{j=1}^{r} P(C_j) G(C_j, \tilde{\mathbf{C}}_{C_j}, P_{C_j}), \end{aligned}$$

where (a) follows from the Postulate Q4 since $\mathbf{C} = \cup_{j=1}^{r} \tilde{\mathbf{C}}_{C_j}$, (b) follows from Postulate Q3 and (c) follows from Postulate Q4. $\square$

Thus we see that Postulates Q3 and Q4 can be regarded as a somewhat stronger version of the *Multiple-choice sequentiality* property expressed by Postulate Q3'.

If we now demand that Postulates Q1 through Q6 hold for the question difficulty function $G(\Omega, \mathbf{C}, P)$ the question is what form this function can possibly take. The answer is given in the following theorem.

Theorem 1 *Let the function $G(\Omega, \mathbf{C}, P)$ where $\mathbf{C} = \{C_1, \ldots, C_r\}$ satisfy Postulates Q1 through Q6. 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*.*

Proof: We can assume, without loss of generality, that there exists a (complete) partition $\mathbf{D} = \{D_1, \ldots, D_N\}$ of Ω into homogeneous subsets D_j, $j = 1, \ldots, N$.

Let $D \subset \Omega$ be a homogeneous subset of the parameter space and let $C \subset D$ be a (free-response) question lying inside of D. Furthermore, let $C' \subset C$ be another question inside of C. Then, according to Postulate Q5,

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

and, since C is homogeneous as well,

$$G(D, C', P_D) = G(D, C, P_D) + G(C, C', P_C). \tag{10}$$

Using the form of $G(\cdot)$ for homogeneous subsets, we obtain from (10)

$$f(P_D(C')) = f(P_D(C)) + f(P_D(C')/P_D(C)),$$

from which it follows, using standard additivity arguments, monotonicity and continuity of the function $f(\cdot)$ (which follow from Postulates Q6 and Q2, respectively) that $f(x) = -c \log x$ where $c > 0$ is a constant (see [37] for details). Since the constant c may depend on the particular homogeneous subset D we can denote it by $u(D)$ and obtain that

$$G(D, C, P_D) = -u(D) \log P_D(C), \tag{11}$$

for any $C \subseteq D$ whenever D is homogeneous.

Substituting (11) into (9) we can obtain

$$G(\Omega, C, P) = G(\Omega, D, P) + f(P(C)/P(D)) = G(\Omega, D, P) - u(D) \log \frac{P(C)}{P(D)},$$

or, equivalently,

$$G(\Omega, C, P) - G(\Omega, D, P) = -u(D) \log P(C) - u(D) \log P(D), \tag{12}$$

where C is an arbitrary subset of D. Then it follows from (12) and continuity of the function G (Postulate Q1) that

$$G(\Omega, C, P) = -u(D) \log P(C) + v(D), \tag{13}$$

for any $C \subseteq D$ whenever D is a homogeneous subset of Ω. Here $v(D)$ is an arbitrary function of D. Setting $P(C) = 1$ in (13) and making use of Postulate Q1, we obtain that $v(D) \equiv 0$ and therefore

$$G(\Omega, C, P) = -u(D) \log P(C). \tag{14}$$

Now let $\mathbf{D} = \{D_1, \ldots, D_N\}$ be a complete partition of Ω into homogeneous subsets D_j, $j = 1, \ldots, N$. Let $C \subset \Omega$ be a free-response question. Then $C = \cup_{j=1}^{N} C \cap D_j$, and since D_j is homogeneous and $C \cap D_j \subseteq D_j$, we obtain using (14) that

$$G(\Omega, C \cap D_j, P) = -u(D_j) \log P(C \cap D_j). \tag{15}$$

On the other hand, by Postulate Q3,

$$G(\Omega, C, P) = G(\Omega, \mathbf{D}_C, P) - G(C, \mathbf{D}_C, P_C), \tag{16}$$

where

$$G(\Omega, \mathbf{D}_C, P) = -\frac{1}{P_C} \sum_{j=1}^{N} u(D_j) P(C \cap D_j) \log P(C \cap D_j), \tag{17}$$

(using the identity $C = \cup_{j=1}^{N} C \cap D_j$, expression (15) and Postulate Q4), and analogously,

$$G(C, \mathbf{D}_C, P_C) = -\sum_{j=1}^{N} u(D_j) \frac{P(C \cap D_j)}{P(C)} \log \frac{P(C \cap D_j)}{P(C)}, \tag{18}$$

Substituting (17) and (18) into (16) we obtain

$$G(\Omega, C, P) = -\sum_{j=1}^{N} \frac{P(C \cap D_j)}{P(C)} u(D_j) \log P(C). \tag{19}$$

We can rewrite (19) as

$$G(\Omega, C, P) = -u(C) P(C) \log P(C), \tag{20}$$

where

$$u(C) \equiv \sum_{j=1}^{N} \frac{P(C \cap D_j) u(D_j)}{P(C)} \tag{21}$$

can be thought of as the definition of function u: $\mathcal{F} \to \mathbb{R}$ for inhomogeneous subsets of Ω. If we define the function $u(\omega)$ on Ω by

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

where $I_D(\omega)$ is the indicator function of a subset $D \subseteq \Omega$, then the expression (21) can be written as

$$u(C) = \frac{\int_C u(\omega) dP(\omega)}{P(C)}. \tag{22}$$

Finally, if $\mathbf{C} = \{C_1, \ldots, C_r\}$ is an arbitrary question, we can use (20) and Postulate Q4 to obtain

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

where the "weights" $u(C_j)$ of the subsets C_j are given by (22). □

Theorem 1 establishes the general form of the question difficulty function if isotropy and linearity conditions are imposed. It appears that the Postulate Q4 (*Mean value*) is the most restricting one of all. It is also the one, as mentioned above, that imposes the linearity constraint on the question difficulty function. The result depends on the measure P and and integrable function u on the parameter space Ω that can be thought of as an attribute of the parameter space. Note that while the measure is extensive, i.e. the measure of a union of two disjoint subsets of Ω is the sum of individual measures ($P(C \cup C') = P(C) + P(C')$ if $C \cap C' = \emptyset$), the function u represents an intensive quantity in that it averages for a union of two disjoint subsets ($u(C \cup C') = \frac{P(C)u(C)+P(C')u(C')}{P(C)+P(C')}$). One can say, loosely speaking, that while measure is similar to volume, u is similar to temperature if physics analogies are to be used. In fact, Appendix describes some insightful parallels between question difficulty on one hand and thermal energy (heat) on the other. These parallels suggest that the function $u(\cdot)$ can be thought of as temperature-like quantity that is allowed to be different at different points of the parameter space. In the following – including follow-up papers [34] and [35] – we refer to the function $u(\omega)$ as *intensity* or *pseudo-temperature*. For the same reason, as mentioned earlier in the paper, it is convenient to think of question difficulty as the amount of *pseudo-energy* associated with the question.

It is also convenient to introduce the *entropy* of question $\mathbf{C}$ as

$$H(\Omega, \mathbf{C}, P) = \frac{\sum_{j=1}^{r} P(C_j) \log \frac{1}{P(C_j)}}{\sum_{j=1}^{r} P(C_j)}, \tag{23}$$

which differs from the pseudo-energy (difficulty) in that it does not involve the pseudo-temperature $u(\cdot)$. It is easy to see that, for any complete question $\mathbf{C} = \{C_1, \ldots, C_r\}$, the expression (23) for question entropy coincides with Shannon entropy of the probability distribution $P(\mathbf{C}) = (P(C_1), \ldots, P(C_r))$ generated by partition $\mathbf{C}$ and measure P on Ω. Moreover, for any complete question $\mathbf{C}$, the pseudo-energy $G(\Omega, \mathbf{C}, P)$ is equal to the *weighted entropy* (studied in [20]) of the same distribution $P(\mathbf{C})$ with the corresponding weights given by the subset pseudo-temperature values $u(C_j)$, $j = 1, \ldots, r$.

It is also easy to see that for a free-response question $C \subset \Omega$, the relationship between pseudo-energy and entropy is simply

$$G(\Omega, C, P) = u(C)H(\Omega, C, P),$$

that is identical to the relationship that exists between thermal energy (heat) and entropy in thermodynamics for reversible processes.

A remark on units of pseudo-energy and pseudo-temperature seems to be in order. It is clear, since the expression for question difficulty is linear in $u(\cdot)$, multiplication of pseudo-temperature function $u(\cdot)$ by any (positive) overall constant would multiply the difficulty of any question by the same constant. A particular choice of this constant corresponds to the choice of units in which pseudo-temperature and pseudo-energy is measured. If just a single information source is considered this choice seems to be largely arbitrary. It appears to be convenient to adapt the convention in which the average pseudo-temperature of the parameter space Ω is equal to 1, i.e. to set the overall scale of $u(\cdot)$ by demanding that $\int_\Omega u(\omega)dP(\omega) = 1$. If two or more information sources need to be compared a different convention turns out to be useful. This issue is discussed further in [35] where information source models are considered.

6 Relationships Between Different Questions

In this section, we assume that all questions are complete (multiple-choice). If $\mathbf{C}'$ and $\mathbf{C}''$ are two arbitrary (complete) questions, the expression $\sum_{C' \in \mathbf{C}'} P(C')G(C', \mathbf{C}''_{C'}, P_{C'})$ will be denoted $G(\Omega, \mathbf{C}''_{\mathbf{C}'}, P)$ and called the *conditional difficulty* of $\mathbf{C}''$. Using this notation, the sequentiality property expressed by Postulate Q3' can be rewritten as

$$G(\Omega, \tilde{\mathbf{C}}, P) = G(\Omega, \mathbf{C}, P) + G(\Omega, \tilde{\mathbf{C}}_{\mathbf{C}}, P), \tag{24}$$

where $\tilde{\mathbf{C}}$ is an arbitrary refinement of $\mathbf{C}$.

If $\mathbf{C}'$ and $\mathbf{C}''$ are two arbitrary (complete) questions and $\mathbf{C} = \mathbf{C}' \cap \mathbf{C}''$ then obviously $\mathbf{C}$ is a refinement of both $\mathbf{C}'$ and $\mathbf{C}''$. One can then write the sequentiality property (24) as

$$G(\Omega, \mathbf{C}, P) = G(\Omega, \mathbf{C}', P) + G(\Omega, \mathbf{C}_{\mathbf{C}'}, P). \tag{25}$$

But it is easy to see that the partition induced by $\mathbf{C} = \mathbf{C}' \cap \mathbf{C}''$ on any set C' in $\mathbf{C}'$ is exactly the same as the partition induced on that set by $\mathbf{C}''$. Therefore, the term $G(\Omega, \mathbf{C}_{\mathbf{C}'}, P)$ in (25) can be equivalently written as $G(\Omega, \mathbf{C}''_{\mathbf{C}'}, P)$ and we arrive at the *chain rule* for the question difficulty which we formulate as a lemma.

Lemma 2 *If $\mathbf{C}'$ and $\mathbf{C}''$ are two arbitrary complete questions and P is a measure on Ω then*

$$G(\Omega, \mathbf{C}' \cap \mathbf{C}'', P) = G(\Omega, \mathbf{C}', P) + G(\Omega, \mathbf{C}''_{\mathbf{C}'}, P).$$

Again, let $\mathbf{C}'$ and $\mathbf{C}''$ be two (complete) questions on Ω and let $\mathbf{C} = \mathbf{C}' \cap \mathbf{C}''$ be the resulting combined question. Then the *pseudo-energy overlap* $J(\Omega, (\mathbf{C}'; \mathbf{C}''), P)$ between $\mathbf{C}'$ and $\mathbf{C}''$ can be defined as the difference between the sum of difficulties of $\mathbf{C}'$ and $\mathbf{C}''$ and that of the combined question $\mathbf{C}' \cap \mathbf{C}''$:

$$J(\Omega, (\mathbf{C}'; \mathbf{C}''), P) = G(\Omega, \mathbf{C}', P) + G(\Omega, \mathbf{C}'', P) - G(\Omega, \mathbf{C}' \cap \mathbf{C}'', P) \tag{26}$$

The definition (26) can be illustrated by a Venn diagram (see Fig. 3). Note that $J(\Omega, (\mathbf{C}'; \mathbf{C}''), P)$ is symmetric with respect to $\mathbf{C}'$ and $\mathbf{C}''$.

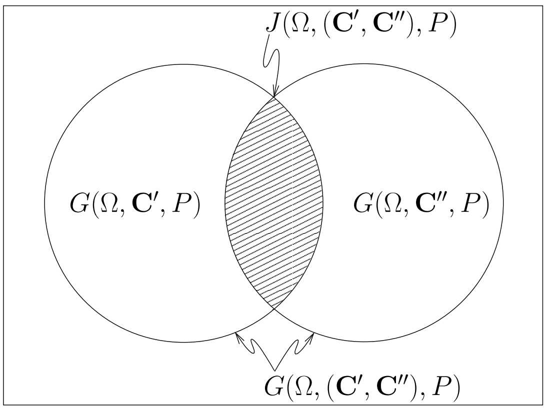

Figure 3: Venn diagram for pseudo-energy overlap.

One can make use of the sequentiality property of pseudo-energy to rewrite expression for the pseudo-energy overlap as follows.

$$\begin{aligned}
J(\Omega, (\mathbf{C}'; \mathbf{C}''), P) &= G(\Omega, \mathbf{C}', P) + G(\Omega, \mathbf{C}'', P) - G(\Omega, (\mathbf{C}', \mathbf{C}''), P) \\
&= G(\Omega, \mathbf{C}', P) + G(\Omega, \mathbf{C}'', P) - G(\Omega, \mathbf{C}'', P) - G(\Omega, \mathbf{C}'_{\mathbf{C}''}, P) \\
&= G(\Omega, \mathbf{C}', P) - G(\Omega, \mathbf{C}'_{\mathbf{C}''}, P).
\end{aligned}$$

We formulate this result as a lemma.

Lemma 3 *If $\mathbf{C}'$ and $\mathbf{C}''$ are two arbitrary questions and P is a measure on Ω then the pseudo-energy overlap $J(\Omega, (\mathbf{C}';\mathbf{C}''), P)$ can be found as*

$$J(\Omega, (\mathbf{C}';\mathbf{C}''), P) = G(\Omega, \mathbf{C}', P) - G(\Omega, \mathbf{C}'_{\mathbf{C}''}, P).$$

Clearly, due to symmetry, the expression for the pseudo-energy overlap stated in Lemma 3 can be equivalently written as $J(\Omega, (\mathbf{C}';\mathbf{C}''), P) = G(\Omega, \mathbf{C}'', P) - G(\Omega, \mathbf{C}''_{\mathbf{C}'}, P)$.

If an expression for the pseudo-energy overlap as a function of the measure P and the difficulty function $u(\omega)$ is desired the definition (26) together with Theorem 1 can be used to obtain

$$J(\Omega, (\mathbf{C}';\mathbf{C}''), P) = \sum_{i=1}^{r'} \sum_{j=1}^{r''} u(C'_i \cap C''_j) P(C'_i \cap C''_j) \log \frac{P(C'_i \cap C''_j)}{P(C'_i)P(C''_j)}. \tag{27}$$

We will be interested in exploring relationships between different questions: given two distinct questions, we would like to know to what degree they are similar to each other. More specifically, if a perfect answer to one question is available, how the difficulty of the other question is affected. To answer this question, let $\mathbf{C}'$ and $\mathbf{C}''$ be two arbitrary complete questions on Ω and let $V^*(\mathbf{C}')$ be a perfect answer to $\mathbf{C}'$. We would like to find an expression for the conditional difficulty of $\mathbf{C}''$ given $V^*(\mathbf{C}')$. Clearly, since a reception of value s'_j of $V(\mathbf{C}')$ updates the measure P to $P_{C'_j}$, the difficulty of $\mathbf{C}''$ given $V(C') = s'_j$ is equal to

$$G(\Omega, \mathbf{C}'', P_{C'_j}) = G(C'_j, \mathbf{C}''_{C'_j}, P_{C'_j}), \tag{28}$$

since subsets of zero measure do not contribute to the difficulty function. Therefore the overall (expected) difficulty $G(\Omega, \mathbf{C}'', V^*(\mathbf{C}'))$ of question $\mathbf{C}''$ given a perfect answer $V^*(\mathbf{C}')$ to $\mathbf{C}'$ can be written as

$$\begin{aligned} G(\Omega, \mathbf{C}'', V^*(\mathbf{C}')) &= \sum_{j=1}^{r'} \Pr(V^*(\mathbf{C}') = s_j) G(\Omega, \mathbf{C}'', P_{C'_j}) \\ &\overset{(a)}{=} \sum_{j=1}^{r'} P(C'_j) G(C'_j, \mathbf{C}''_{C'_j}, P_{C'_j}) = G(\Omega, \mathbf{C}''_{\mathbf{C}'}, P) \\ &\overset{(b)}{=} G(\Omega, \mathbf{C}'', P) - J(\Omega, (\mathbf{C}';\mathbf{C}''), P), \end{aligned} \tag{29}$$

where (a) follows from (28) and the consistency condition (6) – which implies that $\Pr(V^*(\mathbf{C}') = s_j) = P(C_j)$; (b) follows from Lemma 3.

We see from (29) that the conditional difficulty of $\mathbf{C}''$ can be represented as a difference of the standard (unconditional) difficulty and the pseudo-energy overlap $J(\Omega, (\mathbf{C}'; \mathbf{C}''), P)$. Thus the latter provides a measure of reduction of difficulty of a question that is due to a perfect knowledge of an answer to another question. Such a measure can naturally be termed *relative depth* of an answer $V(\mathbf{C}')$ (which in general may not be perfect) with respect to question $\mathbf{C}''$. We can formulate the result just obtained as a lemma.

Lemma 4 *The relative depth of a perfect answer $V^*(\mathbf{C}')$ to question $\mathbf{C}'$ with respect to question $\mathbf{C}''$ is equal to the pseudo-energy overlap between questions $\mathbf{C}'$ and $\mathbf{C}''$.*

The result of Lemma 4 has a clear intuitive interpretation: If two distinct questions are close, i.e. "almost about the same thing" then knowing a (perfect) answer to one of them nearly answers the other one – reduces the difficulty of it to a small value compared to the initial difficulty. The pseudo-energy overlap quantifies the notion of closeness for two arbitrary questions.

7 Examples

We consider an example with a finite parameter space first. Let Ω consist 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). Let all elements be 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. Let $u(GA) = u(GPr) = 1$ reflecting the observation that the green (cold) color is easier to tell from the both yellow and red (warm) colors on one hand, and an apple and a pear are also easy to distinguish from each other because of a different overall shape on the other hand. (Recall that there is no green peach that could be possibly confused with a green apple.) Let $u(YPr) = u(RPr) = 1.5$ reflecting the observation that yellow and red pears can be possibly 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). Finally, let $u(YA) = u(RA) = u(YPc) = u(RPc) = 2$ 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 $\int_\Omega u(\omega)dP(\omega) = 1$ one obtains $u(GA) = u(GPr) = \frac{8}{13}$, $u(YPr) = u(RPr) = \frac{12}{13}$ and $u(YA) = u(RA) = u(YPc) = u(RPc) = \frac{16}{13}$.

The difficulties of free-response questions corresponding to individual elements of Ω can be found as follows: $G(\Omega, GA, P) = G(\Omega, GPr, P) = \frac{8}{13} \cdot \log 8 = \frac{24}{13}$, $G(\Omega, YPr, P) = G(\Omega, RPr, P) = \frac{12}{13} \cdot \log 8 = \frac{36}{13}$ and $G(\Omega, YA, P) = G(\Omega, RA, P) = G(\Omega, YPc, P) = G(\Omega, RPc, P) = \frac{16}{13} \cdot \log 8 = \frac{48}{13}$. The difficulty of the exhaustive multiple choice question (that asks to determine the type and color of the fruit presented to the source) can be found as an expectation of the difficulties of all these free-response questions. Denoting the corresponding (finest) partition of Ω by $\mathbf{C}_f$ we obtain

$$G(\Omega, \mathbf{C}_f, P) = \sum_{\omega \in \Omega} P(\omega) G(\Omega, \omega, P) = 3.$$

Now let us consider difficulties of other multiple-choice questions. Let first of such questions be *"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 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 $P(\overline{C}_g) = \frac{3}{4}$. Thus the difficulty of the question *"Is the fruit green or not?"* can be found as

$$G(\Omega, \{C_g, \overline{C}_g\}, P) = u(C_g) P(C_g) \log \frac{1}{P(C_g)} + u(\overline{C}_g) P(\overline{C}_g) \log \frac{1}{P(\overline{C}_g)} = 0.66$$

Consider another question with subset measures (and thus entropy) equal to those of $\{C_g, \overline{C}_g\}$. Let this question be *"Is the fruit a peach or not?"*. The corresponding partition is $\{C_{Pc}, \overline{C}_{Pc}\}$ 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 $P(\overline{C}_{Pc}) = \frac{3}{4}$. The difficulty of the question $\{C_{Pc}, \overline{C}_{Pc}\}$ is

$$G(\Omega, \{C_{Pc}, \overline{C}_{Pc}\}, P) = u(C_{Pc}) P(C_{Pc}) \log \frac{1}{P(C_{Pc})} + u(\overline{C}_{Pc}) P(\overline{C}_{Pc}) \log \frac{1}{P(\overline{C}_{Pc})} = 0.90$$

We see that this question is somewhat more difficult than the question on whether the fruit is green. The main reason for this difference is that to answer the question on whether the fruit is a peach one might need to have to distinguish a peach from an apple of similar (warm) color which is relatively difficult while answering the question on whether the fruit is green does not involve any "hard" decisions since the color itself is distinct.

Consider now 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 $u(C_g) = \frac{8}{13}$, $u(C_y) = \frac{1}{3} \cdot \frac{12}{13} + \frac{2}{3} \cdot \frac{16}{13} = \frac{44}{39}$, $u(C_g) = u(C_y) = \frac{44}{39}$; $u(C_A) = \frac{1}{3} \cdot \frac{8}{13} + \frac{2}{3} \cdot \frac{16}{13} = \frac{40}{39}$, $u(C_{Pr}) = \frac{1}{3} \cdot \frac{8}{13} + \frac{2}{3} \cdot \frac{12}{13} = \frac{32}{39}$, $u(C_{Pc}) = \frac{16}{13}$. The measures are $P(C_g) = \frac{1}{4}$, $P(C_y) = \frac{3}{8}$, $P(C_r) = \frac{3}{8}$; $P(C_A) = P(C_{Pr}) = \frac{3}{8}$, $P(C_{Pc}) = \frac{1}{4}$. Thus the difficulties of these two questions are

$$G(\Omega, \mathbf{C}_c, P) = u(C_g)P(C_g)\log\frac{1}{P(C_g)} + u(C_y)P(C_y)\log\frac{1}{P(C_y)} + u(C_r)P(C_r)\log\frac{1}{P(C_r)} = \frac{11}{13}\log\frac{8}{3} + \frac{2}{13}\log 4 = 1.51,$$

and

$$G(\Omega, \mathbf{C}_t, P) = u(C_A)P(C_A)\log\frac{1}{P(C_A)} + u(C_{Pr})P(C_{Pr})\log\frac{1}{P(C_{Pr})} + u(C_{Pc})P(C_{Pc})\log\frac{1}{P(C_{Pc})} = \frac{9}{13}\log\frac{8}{3} + \frac{4}{13}\log 4 = 1.60,$$

respectively.

The question about color turns out to be slightly easier than that about type. Qualitatively, the main reason for this difference is that the relatively rare event (that the fruit is green and that it is a peach, respectively) that gives a larger contribution to the difficulty because of the $\log\frac{1}{P(\cdot)}$ factor has smaller average value of pseudo-temperature $u(\cdot)$ in the case of the question about the fruit color.

The pseudo-energy overlap between the "color" and "type" questions can be calculated using the expression (27):

$$J(\Omega, (\mathbf{C}_c; \mathbf{C}_t), P) = \frac{6}{13}\log\frac{4}{3} + \frac{7}{13}\log\frac{8}{9} = 0.100,$$

indicating that while a perfect knowledge of the fruit color helps answering the question about its type, the reduction of difficulty of the "type" question due to the knowledge of color is relatively mild so the question about the fruit type remains almost as hard as it was before the color became known.

For an example with infinite parameter space, consider $\Omega = [0, 1]^2$ with uniform measure P (see Fig. 4 for an illustration). Let $u(\omega) = \frac{3}{2}(\omega_1^2 + \omega_2^2)$ where ω_1 and ω_2 are coordinates on Ω. Let us consider three different questions: $\mathbf{C}_i = \{C_i, \overline{C}_i\}$, where $C_1 = \{\omega : \omega_1 \in [\frac{1}{2}, 1], \omega_2 \in [\frac{1}{2}, 1]\}$, $C_2 = \{\omega : \omega_1 \in [0, \frac{1}{2}], \omega_2 \in [0, \frac{1}{2}]\}$, $C_3 = \{\omega : \omega_1 \in [0, \frac{1}{2}], \omega_2 \in [\frac{1}{2}, 1]\}$. It is easy to see that $P(C_i) = \frac{1}{4}$ for $i = 1, 2, 3$.

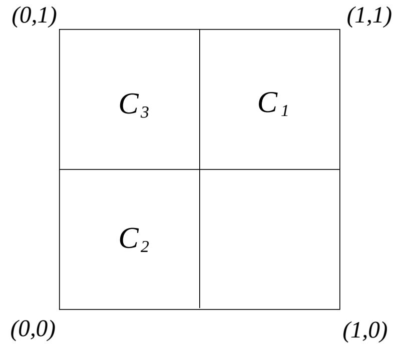

Figure 4: The parameter space $\Omega = [0,1]^2$ and subsets C_i, $i = 1, 2, 3$.

For question $\mathbf{C}_1$, we have $u(C_1) = \frac{3}{2}\int_{\frac{1}{2}}^{1} d\omega_1 \int_{\frac{1}{2}}^{1} d\omega_2(\omega_1^2+\omega_2^2) = \frac{7}{4}$. Then, using the normalization condition $u(C_1)P(C_1) + u(\overline{C}_1)P(\overline{C}_1) = 1$, we can obtain $u(\overline{C}_1) = \frac{3}{4}$, which allows us to compute the difficulty:

$$
\begin{aligned}
G(\Omega, \{C_1, \overline{C}_1\}, P) &= u(C_1)P(C_1)\log\frac{1}{P(C_1)} + u(\overline{C}_1)P(\overline{C}_1)\log\frac{1}{P(\overline{C}_1)} \\
&= \frac{1}{2}\log\frac{4}{3} + \frac{1}{2}\log 4 = 1.208.
\end{aligned}
$$

For question $\mathbf{C}_2$, we obtain $u(C_2) = \frac{3}{2}\int_0^{\frac{1}{2}} d\omega_1 \int_0^{\frac{1}{2}} d\omega_2(\omega_1^2 + \omega_2^2) = \frac{1}{16}$, and, making use of the normalization condition, $u(\overline{C}_2) = \frac{21}{16}$. The difficulty function value for this question becomes

$$
G(\Omega, \{C_2, \overline{C}_2\}, P) = u(C_2)P(C_2)\log\frac{1}{P(C_2)} + u(\overline{C}_2)P(\overline{C}_2)\log\frac{1}{P(\overline{C}_2)} = \log\frac{4}{3} = 0.415.
$$

Finally, for question $\mathbf{C}_3$, we have $u(C_3) = \frac{3}{2}\int_0^{\frac{1}{2}} d\omega_1 \int_{\frac{1}{2}}^{1} d\omega_2(\omega_1^2 + \omega_2^2) = 1$, and, obviously, $u(\overline{C}_3) = 1$. The difficulty function is

$$
\begin{aligned}
G(\Omega, \{C_3, \overline{C}_3\}, P) &= u(C_3)P(C_3)\log\frac{1}{P(C_3)} + u(\overline{C}_3)P(\overline{C}_3)\log\frac{1}{P(\overline{C}_3)} \\
&= \frac{3}{4}\log\frac{4}{3} + \frac{1}{4}\log 4 = 0.811.
\end{aligned}
$$

We see that, among these three questions $\mathbf{C}_1$ turns out to be the most difficult while difficulty of $\mathbf{C}_2$ is the smallest of the three. The reason is that $\mathbf{C}_1$ includes a small measure (rare) set in the region of high values of pseudo-temperature $u(\omega)$. On the other hand, the rare subset in $\mathbf{C}_2$ is located in the region of small values of $u(\omega)$. Question $\mathbf{C}_3$ is naturally placed between these two extremes: its rare subset is located in the region of moderate values of the field $u(\omega)$ so that the difficulty weight of this subset is equal to the average for the whole parameter space.

The overlaps between these questions can easily be computed using expression (27).

$$J(\Omega, (\mathbf{C}_1; \mathbf{C}_2), P) = \frac{1}{2}\log\frac{4}{3} + \frac{1}{2}\log\frac{8}{9} = 0.123,$$

$$J(\Omega, (\mathbf{C}_1; \mathbf{C}_3), P) = \frac{11}{16}\log\frac{4}{3} + \frac{5}{16}\log\frac{8}{9} = 0.232,$$

and

$$J(\Omega, (\mathbf{C}_2; \mathbf{C}_3), P) = \frac{5}{16}\log\frac{4}{3} + \frac{11}{16}\log\frac{8}{9} = 0.013,$$

showing that the most difficult questions – $\mathbf{C}_1$ and $\mathbf{C}_3$ – also exhibit the largest overlap which agrees with the common sense derived notion that knowledge of a perfect answer to a more difficult question can give more help in answering another question.

It is interesting to consider the limit in which the measure of the rare set approaches zero. For this purpose, let $C_1 = \{\omega : \omega_1 \in [1-a, 1], \omega_2 \in [1-a, 1]\}$, $C_2 = \{\omega : \omega_1 \in [0, a], \omega_2 \in [0, a]\}$ and $C_3 = \{\omega : \omega_1 \in [0, a], \omega_2 \in [1-a, 1]\}$ and let $\mathbf{C}_i = \{C_i, \overline{C}_i\}$ for $i = 1, 2, 3$. Let $u(\omega) = \frac{n+1}{2}(\omega_1^n + \omega_2^n)$ where $n \geq 2$ is an integer and $\omega \in \Omega = [0, 1]^2$. Then repeating the calculations for the previously considered example, taking the limit $a \to \infty$ and retaining only terms of the lowest order in a we obtain

$$G(\Omega, \{C_1, \overline{C}_1\}, P) \simeq (n+1)a^2 \log\frac{1}{a} + \log e \cdot a^2 \simeq (n+1)a^2 \log\frac{1}{a},$$

$$G(\Omega, \{C_2, \overline{C}_2\}, P) \simeq \log e \cdot a^2,$$

and

$$G(\Omega, \{C_3, \overline{C}_3\}, P) \simeq 2a^2 \log\frac{1}{a} + \log e \cdot a^2 \simeq 2a^2 \log\frac{1}{a}.$$

Again, we can see that the question $\mathbf{C}_1$ ends up being the most difficult one, with $\mathbf{C}_2$ being the least difficult. It's interesting to note that, to leading order in a, the difficulty of $\mathbf{C}_1$ and $\mathbf{C}_3$ behaves as $a^2 \log \frac{1}{a}$ (with only a numerical coefficient being different), while the difficulty of $\mathbf{C}_2$ behaves as a^2. A related observation is that, in this limit, the difficulty of both $\mathbf{C}_1$ and $\mathbf{C}_3$ is dominated by the rare subset while that of $\mathbf{C}_2$ is dominated by the larger subset with measure approaching 1 since the contribution of the rare subset is diminished by the low value of pseudo-temperature $u(\cdot)$ over that subset.

8 Conclusion

This article initiated development of a quantitative general framework for the description of the process of information extraction from information sources capable of providing answers to given questions. The main motivation for such a framework is the need for optimal decision making in situations characterized with incomplete information and availability of additional information sources. The framework is expected to be especially useful when the knowledge that the information sources possess is of a relatively "loose" variety, i.e. cannot be readily represented in a form admitting direct use in a mathematical formulation. A typical example of such a source would be a human expert who can express a preference for one of the two regions in the parameter space but would find it difficult to produce an accurate probability distribution over the parameter space.

The main components of the proposed framework are questions, answers and information sources. The present article's subject is questions and, in particular, question difficulty functions. The purpose of the latter is measuring the degree of accuracy the given source can achieve on various questions. The idea is that a source would answer easy questions well but its answers' accuracy would decrease with increasing difficulty of questions. The overall form of the question difficulty function is in general determined by the constraints the difficulty function is required to satisfy. The latter constraints depend on the overall properties imposed on the difficulty function. In this article, we assumed the question difficulty to be linear and isotropic on the parameter space. The resulting form was then derived from a system of postulates expressing the desired properties along with more general consistency requirements.

It turns out that the resulting question difficulty function depends on a single scalar quantity $u(\cdot)$ defined on the parameter space and can be interpreted – using parallels with thermodynamics – as an energy-like quantity while the function $u(\cdot)$ takes on the role of temperature that is allowed to take different values at different points of the parameter space. It is interesting to contrast the resulting difficulty function to the corresponding Shannon entropy that is a purely informational quantity measuring the minimum expected number of bits required to communicate a (perfect) answer to the question under consideration. Using parallels with thermodynamics, while the former is similar to thermal energy, the latter can be likened to entropy. It is also worth noting that the linear isotropic model proposed in this article can – in thermodynamics terms – be interpreted as that of *ideal gas*. We expect that other more involved (anisotropic, for instance) versions of question difficulty function would still allow useful interpretations in thermodynamics terms with the difficulty function being similar to thermal energy associated with an appropriate thermodynamical system.

9 APPENDIX

Parallels with Thermodynamics Imagine an ideal gas contained in container A with unit volume $V = 1$ and held at temperature T (see Fig. 5 for an illustration). There is a "marked" molecule. Let $C \subset A$ be a part of the original container that has volume $V < 1$. We are interested in whether the molecule of interest is located in C or otherwise. A "constructive" way of reducing this uncertainty is compressing the original container so that all gas – including the special molecule – is in C with certainty.

The energy conservation law reads $dQ = pdV + dU$, where dQ is the (infinitesimal) heat transferred to the gas, pdV is the work done by the gas and dU is the increment of the gas internal energy. If we insist that the gas be kept at constant temperature T then (since the gas is ideal) $dU = 0$ and the energy conservation law reduces to

$$dQ = pdV. \tag{30}$$

The ideal gas equation of state reads $pV = \nu RT$ where p is the pressure, ν is the amount of substance (in moles) and R is the ideal gas constant. We can use the equation of state to express pressure as a function of the gas volume:

$$p(V) = \frac{\nu RT}{V}. \tag{31}$$

Substituting (31) into (30) we can obtain for the amount of heat transferred to the gas while its volume is educed from 1 to V at constant temperature T:

$$\Delta Q = \int_1^V p(V)dV = \nu RT \int_1^V \frac{dV}{V} = \nu RT \ln V = \nu RT (\ln 2) \log V < 0,$$

implying that the amount of heat equal to

$$-\Delta Q = \ln 2\, \nu RT \log \frac{1}{V} > 0 \tag{32}$$

is taken away from the gas. We can note now that $V = P(C)$ where $P(\cdot)$ is the uniform measure on A describing the initial information on the location of the marked molecule in A. Comparing expression (32) with that for the difficulty of a free-response question C

$$G(\Omega, C, P) = u(C) \log \frac{1}{P(C)}$$

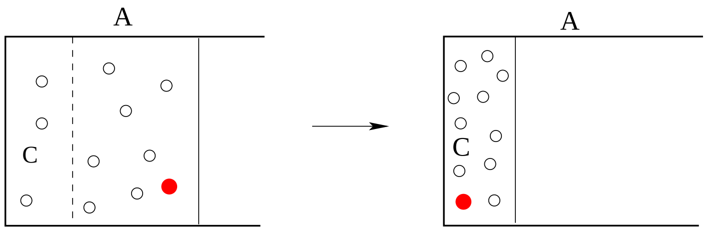

Figure 5: Gas in container A is compressed from original unit volume to volume $V < 1$. The marked molecule is shown as a shaded circle.

we see that (i) the value $u(C)$ plays the role of temperature and (ii) the question difficulty can be thought of as the energy-like quantity that is similar to the thermal energy (heat) that has to be taken away from the system in order to reduce uncertainty about the microstate that can be characterized by entropy. The latter is related to heat by the relationship $dQ = TdS$, where S stands for the thermodynamic entropy. Thus, the higher temperature is the larger the amount of heat that has to be dissipated in order to reduce entropy. Therefore temperature can be interpreted as (thermal) energy per unit of entropy. In application to inquiry, respectively, the pseudo-temperature $u(\cdot)$ can be thought of as the amount of pseudo-energy (difficulty) per unit of Shannon entropy that represents the purely informational quantity measuring the minimum expected number of bits that is necessary to *communicate* a perfect answer to the given question.

*Bibliography

[1] Ben-Tal, A., S. Boyd, A. Nemirovski. 2006. Extending scope of robust optimization: Comprehensive robust counterparts of uncertain problems. *Math. Programming, Ser.B* **107** 63–89.

[2] Ben-Tal, A., A. Nemirovski. 1998. Robust convex optimization. *Math. Oper. Res.* **23**(4) 769–805.

[3] Ben-Tal, A., M. Teboulle. 2007. An old-new concept of convex risk measures: the optimized certainty equivalent. *Mathematical Finance* **17** 449–476.

[4] Birge, J. R., F. Louveaux. 1997. *Introduction to Stochastic Programming*. Springer-Verlag, New York, NY.

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

[6] 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.

[7] 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.

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

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

[10] Cox, R. T. 1946. Probability, frequency, and reasonable expectation. *Am. J. Phys.* **14** 1–13.

[11] Cox, R. T. 1961. *The Algebra of Probable Inference*. Johns Hopkins Press, Baltimore.

[12] Eichhorn, A., W. Römisch. 2005. Polyhedral risk measures in stochastic programming. *SIAM Journal on Optimization* **16** 69–95.

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

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

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

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

[17] 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.

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

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

[20] Guiaşu, Silviu. 1971. Weighted entropy. *Reports on Mathematical Physics* **2**(3) 165–179.

[21] Harmanec, D., G. J. Klir. 1994. Measuring total uncertainty in Dempster-Shafer theory: A novel approach. *Int. J. Gen. Syst.* **22**(4) 405–419.

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

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

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

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

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

[27] 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.

[28] Klir, G. J., D. Harmanec. 1996. Generalized information theory: Recent developments. *Kybernetes* **25**(7/8) 50–66.

[29] Knuth, K. H. 2005. Lattice duality: The origin of probability and entropy. *Neurocomputing* **67** 245–274.

[30] 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.

[31] Maeda, Y., H. Ichihashi. 1993. An uncertainty measure with monotonicity under the random set inclusion. *Int. J. Gen. Syst.* **21**(4) 379–392.

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

[33] 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.

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

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

[36] 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.

[37] 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.

[38] Ruszczyński, A., A. Shapiro. 2006. Optimization of convex risk functions. *Mathematics of Operations Research* **31** 433–452.

[39] Schreiber, T. 2000. Measuring information transfer. *Phys. Rev. Lett.* **85**(2) 461–464.

[40] Shafer, G. 1976. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, NJ.

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

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

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