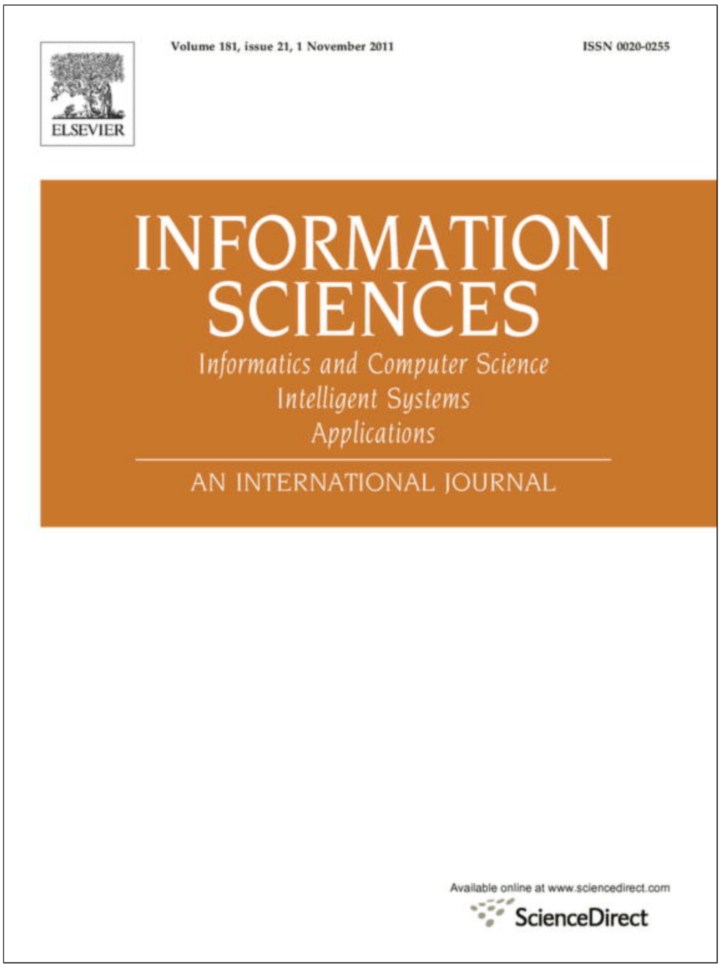
Volume 181, issue 21, 1 November 2011
ISSN 0020-0255
ELSEVIER
INFORMATION SCIENCES
Informatics and Computer Science
Intelligent Systems
Applications
AN INTERNATIONAL JOURNAL
Available online at www.sciencedirect.com
ScienceDirect

Information Sciences 181 (2011) 4847–4859

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Representational information: a new general notion and measure of information

Ronaldo Vigo *

Center for the Advancement of Cognitive Science, Psychology Department, Ohio University, Athens, OH 45701, USA

ARTICLE INFO

Article history:
Received 23 October 2010
Received in revised form 26 January 2011
Accepted 19 May 2011
Available online 27 May 2011

Keywords:
Representational Information
Concepts
Invariance
Complexity
Information measure
Subjective information

ABSTRACT

In what follows, we introduce the notion of representational information (information conveyed by sets of dimensionally defined objects about their superset of origin) as well as an original deterministic mathematical framework for its analysis and measurement. The framework, based in part on categorical invariance theory [30], unifies three key constructs of universal science – invariance, complexity, and information. From this unification we define the amount of information that a well-defined set of objects R carries about its finite superset of origin S, as the rate of change in the structural complexity of S (as determined by its degree of categorical invariance), whenever the objects in R are removed from the set S. The measure captures deterministically the significant role that context and category structure play in determining the relative quantity and quality of subjective information conveyed by particular objects in multi-object stimuli.

1. Introduction

Ever since its inception 63 years ago, Shannon's mathematical theory of communication [23,24] and the information measure on which it is based has significantly shaped the way that engineers, physicists, cognitive scientists, biologists, and statisticians think about information as a measurable quantity [2,3]. Shannon's information measure (SIM) may be traced to the intuition that the more improbable an event is, the more informative it is. To explain, let x be a discrete random variable. How much information is received when we observe a particular value of this variable? Shannon's measure assumes that if a highly probable value for x is detected, then the receiver has gained very little information. Accordingly, if a highly improbable value is detected, the receiver has gained a great amount of information. In other words, the amount of information received from x depends on its probability distribution $p(x)$. SIM is then defined as a monotonic function (i.e., the log function to some base, usually base 2) of the probability of x as follows:

$$h(x) = -\log_2 p(x). \tag{1.1}$$

The idea of linking the probability of an event to its subjective information content as shown in equation 1.1 above has been challenged by many authors, and most recently by Devlin [7] and Luce [19]. The basic criticism being that the measure does not conform to intuitions regarding what humans deem informative. For example, Luce concludes: "*The question remains: Why is information theory not very applicable to psychological problems despite apparent similarities of concepts? ... in my opinion, the most important answer lies in the following incompatibility between psychology and information theory. The elements of choice in information theory are absolutely neutral and lack any internal structure; the probabilities are on a pure, unstructured set whose elements are functionally interchangeable*". The structural component Luce refers to is most clearly manifested by experiments on relational judgments such as those reported in [11,29].

* Tel.: +1 310 292 8778.
E-mail address: vigo@ohio.edu.

0020-0255/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ins.2011.05.020

More telling is the fact that SIM is not consistent with empirical data from experiments that have attempted to establish a connection between information and specific cognitive performance and phenomena. Naturally, these criticisms should be tempered by the fact that the aim of Shannon's theory of communication and of SIM was to characterize information in terms of the transmission and readability of data from the standpoint of electronic systems and not human cognizers [6]. For a genetic history and insight into the aims of Shannon's proposal the reader is referred to [18,22]. However, this fact did not discourage a myriad of social and cognitive researchers from attempting, without success, to use SIM to account for a wide variety of cognitive and behavioral phenomena (the reader is referred to Luce [19] for specific examples). These failures originate from three key restricting aspects of the measure: first, information is operationalized and mediated by ordered sequences of symbols as had been the case in previous models of information such as Hartley's [13]; secondly, the axioms of probability theory (and in particular, the independence axiom) prevent the measure from capturing the impact that context has on events and the subjective nature of probability judgments by human observers; finally, the theory is formulated in terms of event uncertainty (and expectancy) and not in terms of the relational information that is present with certainty in a set of objects.

These limitations appear more severe in view of the fact that there are many situations in the social, psychological, and information sciences where calculating the amount of information conveyed by sets of objects about other sets of objects is important: for example, (1) to determine which subsets of a data set can best summarize its content (under a particular size constraint), (2) to characterize how informative different types of categories (i.e., sets of related objects) and partial categories influence concept learning in human cognizers and machines, (3) to eliminate redundant information from datasets, (4) to compute the prototypical objects of a data set, (5) to account for pattern perception in human and machine cognition, and (6) to measure the amount of contextual information associated with the elements of a set of objects. Note that in each of these situations, relational or contextual information plays a significant role. Thus, in order to develop such a measure, one must depart drastically from the tenets of classical information theory as based on probabilities.

In this paper, we take one such departure by introducing a notion of information that is antithetical to the three restrictive aspects of SIM listed above. *Representational information* (RI) is the information that is conveyed or carried by a particular object or, more generally, a finite non-empty set of objects, say R, about its non-empty superset of origin S, where S is finite and has a certain type of structure (see Fig. 1). By "object" we mean an entity defined in terms of a preset number of dimensions and dimensional values or features. We shall give a more precise definition of these concepts in the following section.

Representational information theory (RIT) is based on five principles: (1) humans communicate via concepts or, in other words, mental representations of categories of objects (where a category is simply a set of objects that are related in some way), (2) therefore, concepts are the mediators of information, not strings, (3) but concepts are relationships between qualitative objects in the environment that are defined dimensionally, (4) the degree of homogeneity of a category (i.e., to what extent its objects are indistinguishable) is characterized quantitatively by its degree of categorical invariance [30,32], and (5) the degree of structural complexity of a category is a function of its degree of categorical invariance and its cardinality (i.e., size) [30,32]. The first three principles are frequently adopted by researchers in the field of human concept learning as may be found in [4,8,11,16,30], while principles four and five form the basis of the theory proposed by Vigo [30,32]. Combined, they support the proposition that the information conveyed by a set of objects R in respect to its superset of origin S is the rate of change in the structural complexity of S whenever the objects in R are removed from S. Thus, the notion of "the degree of surprise of an event" as the basis for measuring information is replaced by the notion of "the rate of change of the structural complexity of a concept" or, equivalently, the rate of change of the structural complexity of the category (in intension) from which the concept is learned.

This idea is in accord with human intuitions about what is informative. Consider that a concept is defined as a mental representation of a category of objects (i.e., a set of objects that are related in some way) [20,30]. Now, if we think of each element R of the power set of S (the power set of S is the set containing all the subsets of S) as a representation of S, then, whenever a receiver receives a representation R of a category S, what is being received is compressed information about the concept associated with the set S. How do we measure the information carried by R about S? If R carries a large amount of information about S, then the absence of its elements from S should either greatly increase or decrease the structural complexity of S. Which means that either S carries a great amount of compressed information as to what R is like or what R is not like. Under this supposition, we propose that the key to measuring representational information (i.e., what R is like according to S) lies in measuring the relative contribution that R makes to the structural complexity of S. If the absence of R (its objects)

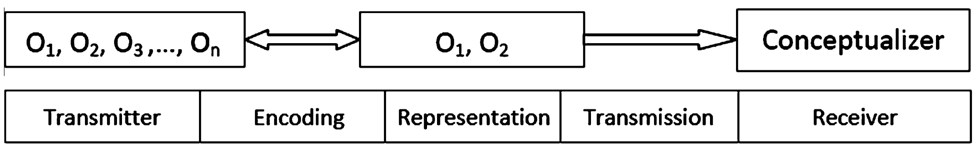

Fig. 1. Schematic of stages in the transmission and reception of representational information. Note that for each environmental categorical stimulus (sets of objects) a concept is acquired by the receiver (conceptualizer) which indicates that the information has been transferred. The condensed representation of the original category (consisting of two objects in the diagram above) is acquired as a partial concept which approximates to various degrees the nature of the original category depending on how much information it carries about it. Our representational information measure determines precisely that amount of information.

in S drastically decreases the complexity of S (and as we shall see, its comprehensibility), then R is not representationally informative about S; if on the other hand, the complexity of S drastically increases without the elements in R, then R is very representationally informative about S. We shall give examples and formalize this idea in the next four sections.

Finally, it should be noted that this quantitative notion of information is abstract and dimensionless. That is, representational information, we posit, is not a qualitative physical property such as length, mass, and Kelvin temperature that can be measured by how much of the qualitative property is present. Rather, representational information is a relative and abstract higher order property measured by the rate of change in the complexity of a stimulus set.

As mentioned, RIT is diametrically opposed to the three general aspects of Shannon's communication theory listed above in that: (1) we replace the notion of a symbolic sequence with that of a concept structure; (2) we base representational information on a deterministic theory known as "categorical invariance theory" and not on probability theory; and lastly, (3) we abandon the notion of an event as a source of information, replacing it instead by sets of objects. In the next section, we shall define in formal terms the notions of a representation, an object, a category, and a concept. In Section 3 we will introduce the reader to categorical invariance theory [30,32]. In Section 4, structural complexity is defined. After these preliminaries, we introduce our measure of representational information in Section 5 as a function of the degree of invariance and structural complexity of a category. Section 6 briefly covers some potential applications of RIT and Section 7 covers some of the limits of RIT and future research directions.

2. Objects, categories, category structures, and representations

A category is a finite set of objects that are related in some way. A "well-defined" category is a category whose content is defined by a Boolean algebraic rule (we shall explain what this means later in this section).[1] Objects are entities defined in terms of a number of preset binary dimensions. Although it may be possible to extend RIT to n-ary valued dimensions ($n \geqslant 2$), this paper only discusses rules involving binary values. Take, for example, a set of objects defined over the three binary dimensions of shape (triangular or circular), color (black or white), and size (large or small). There are $2^3 = 8$ possible objects that may be defined with these three dimensions. An example of one such object is a triangle that is black and small. Using binary variables x, y, and z to represent the dimensions of shape, color, and size respectively, we can represent the object in terms of the conjunctive rule or product rule xyz from Boolean algebra (in other words, the rule "x and y and z") when x = triangular, y = black, and z = small. Furthermore, we can represent the alternative value of a binary dimension x with a superscript prime symbol as in x'. So that, using the variable assignments above, a circle that is black and small may be represented by the Boolean rule $x'yz$. Henceforth, we shall denote binary dimensions with lower case letters of the English alphabet and with such letters accompanied by integer ($n \geqslant 1$) subscripts: e.g., w, x, y, z; $w_1, \ldots, w_n$; $x_1, \ldots, x_n$; $y_1, \ldots, y_n$; $z_1, \ldots, z_n$.

A logical "addition" of such products using the "or" operator (represented by the symbol "+") is known as a Boolean function in DNF or disjunctive normal form (see [14,28] for a discussion of the disjunctive normal form of a Boolean function). Functions in DNF are useful in giving a verbatim description of a category. For example, a category consisting of a triangle that is black and small and a rectangle that is black and small and a rectangle that is white and large may be defined algebraically by the following DNF function, $xyz + x'yz + x'y'z'$, where, once again, the variables with an adjacent prime symbol denote the alternative possible value of the variable (e.g., when x = triangular, x' = rectangular). Because these functions act as membership rules for a category, we shall refer to them as concept functions. In this paper, concept functions will be represented by italicized capital letters of the English alphabet (e.g., F, G, H), while the sets that such functions define in extension will be denoted by function symbols with a set bracket on top. For example, if F is a Boolean function in DNF, $\widehat{F}$ is the category that it defines.

Concept functions are useful in spelling out the logical structure of a category. For example, suppose that x stands for blue, x' stands for red, y stands for circular, and y' stands for square, then the two-variable algebraic expression $x'y + xy$ defines the category consisting of two objects: one "red and round" and the other "blue and round". Clearly, the choice of labels in the expression is arbitrary. Hence, there are many Boolean expressions that define the same category structure and, likewise, different categories with the same structure may be defined by the same Boolean expression. For example, making x' stand for square instead of red and making y stand for blue instead of round defines the structurally equivalent category consisting of a "blue square" object and a "blue circular" object, where the relationships between the dimensional values remain the same. These structurally equivalent categories form category types (or distinct structures) and may be represented by a single concept function. A class of well-defined category types whose category instances are defined by D dimensions and contain p objects is called a $D[p]$ family. For instance, the well-defined category described above belongs to the 2[2] family since it is comprised of two objects that are defined by two dimensions (color and shape). For a catalog of these category structures see [1,9,13].

Category structures such as those cataloged in [9] have been studied from the standpoint of the concepts they convey to humans (for a non-technical introduction to concepts and the connection between categories and concepts the reader is referred to [20,31]). Indeed, one of the primary goals of theories of concept learning is to be able to predict the degree of difficulty experienced by humans when learning concepts from different types of category structures (see Fig. 2 below for examples of instances of category structures). A typical experimental paradigm used to determine the degree of learning difficulty of an instance of a category structure consists of showing the members of the category along with the members

[1] By convention, we shall say that the empty set is also a well-defined category.

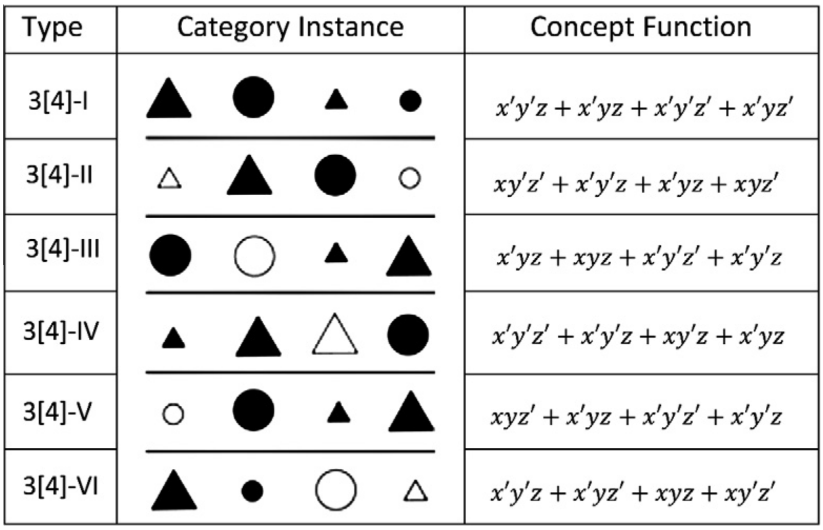

Fig. 2. Instances of the 3[4] category types where *x* represents the color dimension, *y* represents the shape dimension, and *z* represents the size dimension.

of the complement of the category (the set of objects definable by the *D* dimensions that are not in the category) one at a time for a certain period of time. For each object presented, the human subject presses one of two buttons indicating whether the object belongs or does not belong to the category, after which a prompt is shown indicating whether or not the object was classified correctly: the presumption being that if the subject has acquired the concept associated with the category, she should be able to classify all the objects correctly. After the subject reaches a performance criterion (e.g., 100% correct), the next block of trials begins in order to test another instance of the same or of another category structure.

A second way of testing concept learning performance involves a training phase where subjects are shown the entire instance of a category structure (i.e., all the objects of the category simultaneously as opposed to one at a time) for a certain amount of time. Subjects are then asked to classify objects as described above but without corrective feedback. Under the first protocol, degree of concept learning difficulty is operationalized by the number of trials necessary to reach the correctness criterion. Under the second protocol, degree of learning difficulty is operationalized by the percentage of classification errors.

One of the most important families of category structures that has been studied empirically using both of these approaches (and others) is the 3[4] family consisting of six category structures or types (for a proof that there are precisely six distinct structures in this family, see [1,15]). In a now classic experiment, Shepard et al. [25] observed the following increasing learning difficulty ordering in the part of human subjects: I < II < [III, IV, V] < VI (with types III, IV, and V of approximately the same degree of difficulty). Fig. 2 illustrates visual instances of the 3[4] family structure types in the form of simple geometric shapes. This 3[4] family ordering has been empirically replicated numerous times by several researchers [17,21,25,32].

Since the ultimate goal of this paper is to characterize representational information in terms of the transmission of concepts as mediated by category structures, it is important to appreciate the robustness of this and other empirical results like it. Vigo [30,32] introduced the first general invariance principle able to account successfully for this and other similar results. Based on this principle, Vigo [30,32] introduces a candidate mathematical law of invariance that makes accurate predictions about the degree of concept learning difficulty of the structures in the 3[4] family and, in principle, any well-defined category structure. The law has been verified empirically for a wide range of category structures and links degree of concept learning difficulty to the structural complexity of a category. That is, the structural complexity of any category as defined by Vigo [30,32] is consistent with empirical findings on the degree of difficulty experienced by humans while apprehending concepts from well-defined categories.

Before embarking on the details of the representational information measure, we shall first define the notion of a representation (or "representative") of a well-defined category. A representation of a well-defined category S is any subset of S. The power set $\wp(S)$ is the set of all such representations. Since there are $2^{|S|}$ elements in $\wp(S)$, then there are $2^{|S|}$ possible representations of S ($|S|$ stands for cardinality or size of the set S). Some representations capture the structural (i.e., relational) "essence" or nature of S better than others. In other words, some representations carry more representational information (i.e., more conceptual significance) about S than others. For example, consider a well-defined category with three objects defined over three dimensions (color, shape, and size) consisting of a small black circle, a small black triangle, and a large white circle. The small black circle better captures the character of the category as a whole than does the large white circle.

In addition, it would seem that: (1) for any well-defined category S, all the information in S is conveyed by S itself, and that (2) the empty set ϕ carries no information about S. The aim of our measure is to measure the amount and quality of conceptual information carried by representations or representatives of the category S about S while obeying these two basic requirements and capturing the conceptual significance of S.

3. Categorical invariance

Vigo [30,32] defines categorical invariance as the high-order property of a well-defined category to stay the same (in respect to its object content) after its objects are transformed relative to each of their defining dimensions. To illustrate this idea, consider the category containing a triangle that is black and small, a circle that is black and small, and a circle that is white and large. This category is described by the concept function $xyz + x'yz + x'y'z'$. Let's encode the features of the objects in this category using the digits "1" and "0" so that each object may be representable by a binary string. For example, "111" stands for the first object when $x = 1 = triangular$, $y = 1 = small$, and $z = 1 = black$. Thus, the entire set can be encoded by {111, 011, 000}. If we transform this set in terms of the shape dimension by assigning the opposite shape value to each of the objects in the set, we get the perturbed set {011, 111, 100}. Now, if we compare the original set to the perturbed set, they have two objects in common with respect to the dimension of shape. Thus, two out of three objects remain the same. This is a measure of the partial invariance of the category with respect to the dimension of shape. The first pane of Fig. 3 illustrates this transformation. Doing this for each of the dimensions, we can form an ordered set, or vector, of values consisting of these partial invariants which we refer to as the *structural* or *logical manifold* of the concept function or category type (see Fig. 3).

Formally, these partial invariants can be represented in terms of a vector of discrete partial derivatives of the concept function that defines the Boolean category. This is shown in Eq. (3.3) below where $\Lambda(F)$ stands for the logical manifold of the concept function F and where a "hat" symbol over the partial differentiation symbol indicates discrete differentiation (for a detailed and rigorous explanation, see [30]).

Discrete partial derivatives are completely analogous to continuous partial derivatives in Calculus. Loosely speaking, in Calculus, the partial derivative of an n variable differentiable (and therefore continuous) function $f(x_1, \ldots, x_n)$ is defined as how much the function value changes relative to how much the input value(s) change as seen below:

$$\frac{\partial f(x_1,\ldots,x_n)}{\partial x_i} = \lim_{\Delta x_i \to 0} \frac{f(x_1 \ldots, x_i + \Delta x_i, \ldots x_n) - f(x_1,\ldots,x_n)}{(x_i + \Delta x_i) - x_i}. \tag{3.1}$$

On the other hand, the discrete partial derivative, defined by the equation below (where $x_i' = 1 - x_i$ with $x_i \in \{0,1\}$) is totally analogous to the continuous partial derivative except that there is no limit taken because the values of x_i can be only 0 or 1:

$$\frac{\hat{\partial} F(x_1,\ldots,x_n)}{\hat{\partial} x_i} = \frac{F\left(x_1 \ldots, x_i', \ldots x_n\right) - F(x_1,\ldots,x_n)}{x_i' - x_i}. \tag{3.2}$$

The value of the derivative is ±1 if the function assignment changes when x_i changes, and the value of the derivative is 0 if the function assignment does not change when x_i changes. Note that the value of the derivative depends on the entire vector $(x_1, \ldots, x_n)$ (abbreviated as $\vec{x}$ in this article) and not just on x_i. As an example, consider the concept function AND, denoted as $F(\vec{x}) = F(x_1, x_2) = x_1 x_2$ (Equivalently, we could also write this function as in the examples under Section 2 as $F(x, y) = xy$. Because this is more readable than the vector notation, we shall continue using it in other examples.) Also, consider the particular point $\vec{x} = (0, 0)$. At that point, the derivative of the concept function AND with respect to x_1 is 0 because the value of the concept function does not change when the stimulus changes from $(0, 0)$ to $(1, 0)$. If instead we consider the point$(0, 1)$, the derivative of AND with respect to x_1 is 1 because the value of the concept function does change when the stimulus changes from $(0, 1)$ to $(1, 1)$.

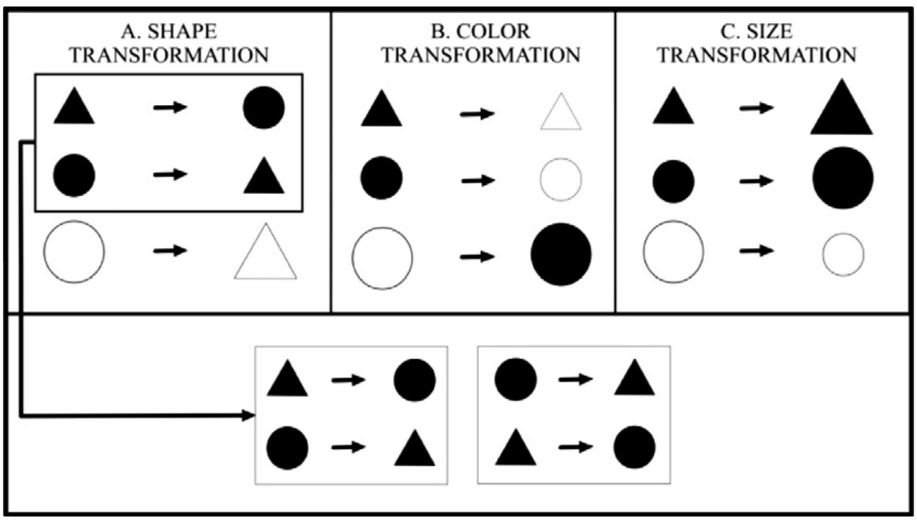

Fig. 3. Logical manifold transformations along the dimensions of shape, color, and size for an instance of a category type from the 3[3] family. The fourth pane underneath the three top panes contains the pairwise symmetries revealed by the shape transformation.

Accordingly, the discrete partial derivatives in Eq. (3.3) below determine whether a change has occurred in the category in respect to a change in each of its dimensions. The double lines around the discrete partial derivatives give the proportion of objects that have not changed in the category and are defined in Eq. (3.4) below:

$$\Lambda(F) = \left(\left\| \frac{\hat{\partial} F(x_1, \ldots, x_D)}{\hat{\partial} x_1} \right\|, \left\| \frac{\hat{\partial} F(x_1, \ldots, x_D)}{\hat{\partial} x_2} \right\|, \ldots, \left\| \frac{\hat{\partial} F(x_1, \ldots, x_D)}{\hat{\partial} x_D} \right\| \right). \tag{3.3}$$

$$\Lambda_i(F) = \left\| \frac{\hat{\partial} F(x_1, \ldots, x_D)}{\hat{\partial} x_i} \right\| = 1 - \left[\frac{1}{p} \sum_{\vec{x}_j \in \widehat{F}} \left| \frac{\hat{\partial} F(\vec{x}_j)}{\hat{\partial} x_i} \right| \right]. \tag{3.4}$$

In the above definition (Eq. (3.4)), $\vec{x}$ stands for an object defined by a vector of D dimensional values $(x_1, \ldots, x_D)$. The general summation symbol represents the sum of the partial derivatives evaluated at each object $\vec{x}_j$ from the Boolean category $\widehat{F}$ (this is the category defined by the concept function F). The partial derivative transforms each object $\vec{x}_j$ in respect to its ith dimension and evaluates to 0 if, after the transformation, the object is still in $\widehat{F}$ (it evaluates to 1 otherwise). Thus, to compute the proportion of objects that remain in $\widehat{F}$ after changing the value of their ith dimension, we need to divide the sum of the partial derivatives evaluated at each object $\vec{x}_j$ by p (the number of objects in $\widehat{F}$) and subtract the result from 1. The absolute value symbol is placed around the partial derivative to avoid a value of negative 1 (for a detailed explanation, see [30]).

Relative degrees of total invariance across category types from different families can then be measured by taking the Euclidean distance of each structural or logical manifold (Eq. (3.3)) from the zero logical manifold whose components are all zeros (i.e., $(0, \ldots, 0)$). Thus, the overall degree of invariance Φ of the concept function F (and of any category it defines) is given by the equation below:

$$\Phi(F(x_1, \ldots, x_D)) = \left[\sum_{i=1}^{D} \left[\left\| \frac{\hat{\partial} F(x_1, \ldots, x_D)}{\hat{\partial} x_i} \right\| \right]^2 \right]^{1/2}. \tag{3.5}$$

Using our example from pane one in Fig. 3, we showed that the original category and the perturbed category have two elements in common (out of the three transformed elements) in respect to the shape dimension; thus, its degree of partial invariance is expressed by the ratio 2/3. Conducting a similar analysis in respect to the dimensions of color and size, its logical manifold computes to $\left(\frac{2}{3}, \frac{0}{3}, \frac{0}{3}\right)$ and its degree of categorical invariance is

$$\Phi\left(x_1x_2x_3 + x_1'x_2x_3 + x_1'x_2'x_3'\right) = \sqrt{\left(\frac{2}{3}\right)^2 + \left(\frac{0}{3}\right)^2 + \left(\frac{0}{3}\right)^2} \approx .67. \tag{3.6}$$

Note that the concept function $xyz + x'yz + x'y'z'$ used in our example at the beginning of Section 3 has been rewritten in an entirely equivalently form as $x_1x_2x_3 + x_1'x_2x_3 + x_1'x_2'x_3'$ in order to be consistent with the vector notation introduced above. Henceforth, we shall use both ways of specifying concept functions and it will be left to the reader to make the appropriate translation. We do this since the non-vector notation is more intuitive and less confusing to comprehend structurally.

Invariance properties facilitate concept learning and identification. More specifically, the proposed mathematical framework reveals the pairwise symmetries that are inherent to a category structure when transformed by a change to one of its defining dimensions. One such pairwise symmetry is illustrated in the bottom pane of Fig. 3. The more of these symmetries, the less the dimension is useful in determining category membership. In others words, the dimensions associated with high invariance do not help us discriminate the perturbed objects from the original objects in terms of category membership. Consequently, these particular dimensions do not carry "diagnostic" information about their associated category; however, they signal the presence of redundant information.

This redundant information is useful because it identifies first hand which aspects of the categorical stimulus could be ignored while learning the concept and while attempting to construct membership rules. Conversely, the lower the degree of invariance of a category, the more impact its dimensions have on the character of the category as a whole due to their mutual interdependency and the more mutually distinguishable the members of the category are. In short, the degree of categorical invariance of a category may be construed as either a measure of the degree of mutual distinctiveness or, more directly, as a measure of the overall degree of likeness (homogeneity) between the elements of the category. This property can be easily seen when comparing category type 1 to category type 6 in Fig. 2. Note that any possible pattern between dimensional values or features in type 6 is broken up by the way that these values are distributed throughout the objects without any particular one being dominant.

Finally, we should mention that the categorical invariance principle is *not equivalent* to factoring out variables from the concept function formulae (in disjunctive normal form) that define the category structure. To recognize this, note that the variables of the sixth and last concept function in the table (i.e., type 3[4]–VI) in Fig. 2 may be factored out in several ways: yet, the degree of invariance of the concept function is zero. In principle, there are an infinite number of similar examples that demonstrate this important fact.

4. Structural complexity

Using the definition of categorical invariance (Eq. (3.5) above), we define the structural complexity Ψ of a well-defined category $\widehat{F}$ and its subjective structural complexity ψ. The structural complexity of a well-defined category $\widehat{F}$ is directly proportional to the cardinality of the category or size (i.e., the number of elements in the category) and indirectly proportional to a monotonically increasing function f of the degree of invariance of the concept function F that defines the category. This relationship is expressed formally by Eq. (4.1) below. The intuition here is that the raw complexity measured by the number of items in the category is cut down or diminished by the degree of set homogeneity or patternfullness of the category as measured by its degree of invariance (see Section 3 for an explanation of categorical invariance and set homogeneity):

$$\Psi\left(\widehat{F}\right) = \frac{p}{f(\Phi(F))}. \tag{4.1}$$

The simplest function that meets the above criterion is the identity function. Thus, we use it as a baseline standard to define the structural complexity of a category. Moreover, since the degree of categorical invariance Φ of the concept function F can potentially be equal to zero, we have added a 1 to it to avoid division by zero in 4.1 above. Then, the structural complexity Ψ of a category $\widehat{F}$ is directly proportional to its cardinality and indirectly proportional to its degree of invariance (plus one):

$$\Psi\left(\widehat{F}\right) = \frac{p}{\Phi(F)+1} = \frac{p}{\left[\sum_{i=1}^{D}\left[\left\|\frac{\hat{\partial} F(x_1,\ldots,x_D)}{\hat{\partial} x_i}\right\|\right]^2\right]^{\frac{1}{2}} + 1}. \tag{4.2}$$

Although 4.2 above is a good predictor of the subjective structural complexity of a well-defined category (as indicated by how difficult it is to apprehend it) [30,32], it has been shown empirically that subjective structural complexity judgments may more accurately obey an exponentially decreasing function of its degree of invariance [30,32]. Thus, we define the subjective structural complexity ψ of a category $\widehat{F}$ as being directly proportional to its cardinality and indirectly proportional to the exponent of its degree of invariance:

$$\psi\left(\widehat{F}\right) = pe^{-\Phi(F)} = pe^{-\left[\sum_{i=1}^{D}\left[\left\|\frac{\hat{\partial} F(x_1,\ldots,x_D)}{\hat{\partial} x_i}\right\|\right]^2\right]^{1/2}}. \tag{4.3}$$

There are parameterized variants of Eqs. (4.2) and (4.3) above with cognitively motivated parameters [30,32].[2] The parameterized versions of the measures, while less parsimonious, do account for individual differences in the perception of the structural complexity of a well-defined category and, consequently, the subjective degree of difficulty experienced by human observers when acquiring concepts from their corresponding well-defined categories.

5. Representational information

With the preliminary apparatus introduced in Sections 2–4, we are now in a position to introduce a measure of representational information that meets the goals set forth in the introduction to this paper. In general, a set of objects is informative about a category whenever the removal of its elements from the category increases or decreases the structural complexity of the category as a whole. That is, the amount of representational information (RI) conveyed by a representation R of a well-defined category $\widehat{F}$ is the rate of change of the structural complexity of $\widehat{F}$. Simply stated, the representational information carried by an object or objects from a well-defined category $\widehat{F}$ is the percentage increase or decrease (i.e., rate of change or growth rate) in structural complexity that the category experiences whenever the object or objects are removed.[3]

More specifically, let $\widehat{F}$ be a well-defined category defined by the concept function F and let the well-defined category R be a representation of $\widehat{F}$ (i.e., $R \subseteq \widehat{F}$ or $R \in \wp(\widehat{F})$). Then, if $\widehat{G} = \widehat{F} - R$, the amount of representational information h_o of R in respect to $\widehat{F}$ is determined by Eq. (5.1) below where $|\widehat{F}|$ and $|\widehat{G}|$ stand for the number of elements in $\widehat{F}$ and in $\widehat{G}$, respectively:

[2] For the readers' convenience, the parameterized variant of Eq. (4.3) as introduced in [32] follows: $\psi\left(\widehat{F}\right) = pe^{-k\left[\sum_{i=1}^{D}\left[\alpha_i\left[\left\|\frac{\hat{\partial} F(x_1,\ldots,x_D)}{\hat{\partial} x_i}\right\|_c\right]\right]^s\right]^{\frac{1}{s}}}$. The parameter α_i stands for a human observer's degree of sensitivity to (i.e., extent of detection of) the invariance pattern associated with the ith dimension (this is usually a number in the closed real interval [0, 1]). k is a scaling parameter (a real number greater than 0) that indicates the overall ability of the subject to discriminate between mental representations of logical manifolds known as ideotypes (a larger number indicates higher discrimination) and c is a constant parameter in the closed interval [0, 1] which captures possible biases displayed by observers toward invariant information (c is added to the numerator and the denominator of the ratios that make up the logical or structural manifold of the well-defined category). Finally, s is a parameter that indicates the most appropriate measure of distance as defined by the generalized Euclidean metric (i.e., the Minkowski distance measure). In our investigation [30,32], the best predictions are achieved when $s = 2$ (i.e., when using the Euclidean metric). Please note that it has been shown that using the scaling parameter k alone in equation (4.3) (without the help of any other of the aforementioned parameters) enables (4.3) to account for over 95% of the variance in the human concept learning data. Optimal estimates of these free parameters on the aggregate data provide a baseline to assess any individual differences encountered in the pattern perception stage of the concept learning process and may provide a basis for more accurate measurements of subjective representational information.

[3] We could simply define the representational information of a well-defined category as the derivative of its structural complexity. We do not because our characterization of the degree of invariance of a concept function is based on a discrete counterpart to the notion of a derivative in the first place.

$$\hbar_0\left(\mathrm{R}\middle|\widehat{F}\right) = \frac{\Psi\left(\widehat{G}\right) - \Psi\left(\widehat{F}\right)}{\Psi\left(\widehat{F}\right)} = \frac{\left(\frac{\left|\widehat{G}\right|}{\Phi(G(\vec{x}))+1}\right) - \left(\frac{\left|\widehat{F}\right|}{\Phi(F(\vec{x}))+1}\right)}{\left(\frac{\left|\widehat{F}\right|}{\Phi(F(\vec{x}))+1}\right)}. \tag{5.1}$$

Likewise, let $\widehat{F}$ be a well-defined category defined by the concept function F and let the well-defined category R be a representation of $\widehat{F}$ (i.e., $\mathrm{R} \subseteq \widehat{F}$ or $\mathrm{R} \in \wp(\widehat{F})$). Then, if $\widehat{G} = \widehat{F} - \mathrm{R}$, the amount of subjective representational information $\hbar_s$ of R in respect to $\widehat{F}$ is determined by Eq. (5.2) below:

$$\hbar_s\left(\mathrm{R}\middle|\widehat{F}\right) = \frac{\psi\left(\widehat{G}\right) - \psi\left(\widehat{F}\right)}{\psi\left(\widehat{F}\right)} = \frac{\left|\widehat{G}\right| \cdot e^{-\Phi(G(\vec{x}))} - \left|\widehat{F}\right| \cdot e^{-\Phi(F(\vec{x}))}}{\left|\widehat{F}\right| \cdot e^{-\Phi(F(\vec{x}))}}. \tag{5.2}$$

Note that definitions 5.1 and 5.2 above yield negative and positive percentages. Negative percentages represent a drop in complexity, while positive percentages represent an increase in complexity. Thus, RI has two components: a magnitude and a direction (just as the value of the slope of a line indicates both magnitude and direction). For humans, the direction of RI is critical: for example, a relatively large negative value obtained from 5.1 and 5.2 above indicates a small amount of RI, while a relatively large positive value indicates a large amount of RI. In the following examples, it will be shown that, intuitively, the RI values make perfect sense for representations of the same size (i.e., with the same number of objects). In fact, it should be noted that, in general, it is not meaningful to compare RI values of representations of different cardinality or size if one wishes to cancel out the effect that category size alone could have on complexity reduction. Indeed, it should be noted that when comparing RI values for representations of different sizes, there will be cases when a smaller representation (e.g., a singleton) may in fact be more representationally informative than a representation with a greater number of objects. This is due to the fact that the smaller representation informs us more efficiently about the structural pattern underlying the category of origin or source category $\widehat{F}$. In other words, representational information as a whole tells us which representations are *relatively better at compressing* the relational and qualitative information in $\widehat{F}$.

And this is consistent with the fact that the measure above measures both: (1) the *magnitude* of relational information in S that is captured by its representation (as the rate of change in structural complexity), and (2) the *quality* (or goodness) of the representation (as the direction of that rate of change). Another way of interpreting this duality is by thinking of RI as measuring *goodness of representation* in terms of the percentage of relational information captured by the representation (i.e., how faithful is the representation to its source), and in terms of the parsimony or minimality of the representation in respect to its source. However, contrary to this dual character of the measure, multiple object representations may be fallaciously perceived as more informative than single object representations if considered on the basis of size alone. Thus, when running human experiments on subjective information judgments, experimenters should take care that these two aspects of the measure are understood and differentiated. On the other hand, if the user is only interested in the magnitude of RI, then taking the absolute value of 5.1 and 5.2 above will yield the desired values.

Finally, note that the measure yields zero percent information only when R is the empty set. This is consistent with the assumption that the empty set does not convey relational information about its category of origin. On the other hand, when R is $\widehat{F}$ itself, it carries all the relational information possible about $\widehat{F}$, and as anticipated, its RI value in magnitude alone is 100% or 1; however, this value is negative since there is a decrease in complexity. As such, it conveys little RI. This makes sense because when a category represents itself, there is no compression at all, and therefore, no simplification. Thus, a bad representation or bad "relational information compressor" of a well-defined category $\widehat{F}$, when quality and magnitude are both considered, is $\widehat{F}$ itself.

Using Eq. (5.2) above, we can compute the amount of subjective representational information associated with each representation of any category instance defined by any concept function. Take the category defined by the concept function $xyz + x'yz + x'y'z'$ where x = triangular, y = black, and z = small used in Section 3 as an example (Fig. 5 displays the category). To be consistent with the vector notation introduced under Section 3, this concept function can also be written as: $x_1x_2x_3 + x_1'x_2x_3 + x_1'x_2'x_3'$, and as before, we leave it up to the reader to make the necessary translation. As under Section 3, the objects of this category may be encoded in terms of zeros and ones, and the category may be encoded by the set {111, 011, 000} to facilitate reference to the actual objects. The amount of subjective representational information conveyed by the singleton (single element) set containing the object encoded by 111 (and defined by the rule xyz) in respect to the category encoded by {111, 011, 000} (and defined by the concept function $xyz + x'yz + x'y'z'$')) is computed as shown in 5.3 and 5.4 below:

$$\hbar_s(\{111\}|\{111,011,000\}) = \frac{\psi\left(\widehat{G}\right) - \psi\left(\widehat{F}\right)}{\psi\left(\widehat{F}\right)} = \frac{\left|\widehat{G}\right| \cdot e^{-\Phi(x'yz + x'y'z')} - \left|\widehat{F}\right| \cdot e^{-\Phi(xyz + x'yz + x'y'z')}}{\left|\widehat{F}\right| \cdot e^{-\Phi(xyz + x'yz + x'y'z')}} \tag{5.3}$$

Next, we use compute the value of $\Phi(xyz + x'yz + x'y'z')$ as in example 3.6 above and compute $\Phi(x'yz + x'y'z')$ using its logical or structural manifold (0,0,0) and get:

R	$\widehat{F}$	$\widehat{G} = \widehat{F} - R$	$\hbar_s(R\mid\widehat{F})$
{111}	{111, 011, 000}	{011, 000}	.30
{011}	{111, 011, 000}	{111, 000}	.30
{000}	{111, 011, 000}	{111, 011}	-.52

Fig. 4. Amount of information conveyed by all the possible single element representations of $\widehat{F}$.

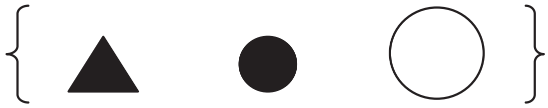

Fig. 5. Category instance of $xyz + x'yz + x'y'z'$ concept function.

$$\hbar_s(\{111\}|\{111,011,000\}) = \frac{\left|\widehat{G}\right| \cdot e^{-\Phi(x'yz + x'y'z')} - \left|\widehat{F}\right| \cdot e^{-\Phi(xyz + x'yz + x'y'z')}}{\left|\widehat{F}\right| \cdot e^{-\Phi(xyz + x'yz + x'y'z')}} = \frac{2e^{-0} - 3e^{-.67}}{3e^{-.67}} \approx \frac{2 - 3 \cdot .51}{3 \cdot .51} \approx .30 \tag{5.4}$$

Similarly, if we compute the results for the remaining two singleton (single element) representations of the set {111, 011, 000}, we get the values shown in the table of Fig. 4 above. These illustrate that the representation {000} is relatively less informative in respect to its category of origin {111, 011, 000} since the absence of 000 results in a 52% reduction in the structural complexity of $\widehat{F}$ (i.e., $-.52$). Likewise, the other two singleton representations are more informative since the presence of their elements (111 and 011, respectively) in $\widehat{F}$ increase the structural complexity of $\widehat{F}$ by 30%. The reader is directed to Fig. 5 above, showing a visuo-perceptual instance of the category structure, in order to confirm these results intuitively.

Fig. 6 shows the information conveyed by the single element representations of the six category structures in the 3[4] family of structures (see Fig. 2 for visuo-perceptual instances of each type). Information vectors containing the amount of information conveyed by each single object representation are given in the information column. Note that each of the single element representations of category structures 3[4]–1, 3[4]–2, and 3[4]–6, respectively convey the same amount of information.

As mentioned under Section 1, it should be noted that this quantitative notion of information is a relative, abstract, and dimensionless. That is, representational information, we posit, is not a qualitative physical property such as length, mass, and Kelvin temperature that can be measure by how much of the qualitative property is present. Rather, representational information is a relative and abstract higher order property measured by the rate of change in the complexity of the structure of the stimulus set.

Furthermore, in general, representational information is non-additive at the level of singleton representations (i.e., representations with a single object). In other words, since the total representational information conveyed by a well-defined category is simply a percentage computed in respect to itself, this value may not be computed by simply adding the representational information of each singleton subset of the category. The reason is that each singleton representation generates a different context that bears a different quantitative relation to its set of origin S: thereby generating unsystematic percentages. That this is the case is clear from Fig. 6. If RI were additive at the level of singleton representations, it would have to be the case that for each of the six well-defined categories listed, the sum of the RI of its singleton representations would be equal to 1, which is clearly not the case.

Although, under particular constraints, it may be the case that an additive property would hold for particular kinds of representations or subsets of a well-defined category, this is a matter that will require further investigation. Indeed, we believe that it is this resistance to an easy additive characterization that makes the RI measure extremely effective in capturing contextual effects. As such, it would seem that structural sensitivity comes at some cost and that, if a highly constraint additive property were possible, it would be under conditions involving a high degree of contextual independency between the objects in some of the representations of the well-defined category.

Finally, note that singleton representations that are considered in *isolation* (i.e., without a category of origin from which they get their representational meaning) as well-defined categories, only have two representations (the empty set and the full set). However, the same singleton categories acting as representations of a well-defined category S, may be regarded as the minimal size representations of S, or as the representational atoms of S. If these representational atoms form the base of a pyramid-like structure made up of the representations of S, then, at the very top of the pyramid is the maximal representation of S containing $|S|$ objects.

Category	Objects	Information
3[4]-1	{001, 011, 000, 010}	[.20, .20, .20, .20]
3[4]-2	{100, 001, 011, 110}	[.05, .05, .05, .05]
3[4]-3	{011, 111, 000, 001}	[-.08, -.31, -.31, -.08]
3[4]-4	{000, 001, 101, 011}	[-.31, .78, -.31, -.31]
3[4]-5	{110, 011, 000, 001}	[-.41, -.22, -.22, .52]
3[4]-6	{001, 010, 111, 100}	[-.25, -.25, -.25, -.25]

Fig. 6. Amount of information conveyed by all the possible single element representations of six different category types or concept functions.

6. Potential applications

As enunciated in the abstract of this paper, the aim of RIT is to provide an alternative mathematical account of the nature of information that may contribute to overcoming the practical and conceptual failures of SIT in cognitive research. Thus, RIT introduces an "in principle" approach to overcoming the limitations of SIT. However, beyond providing an alternative theoretical way of understanding and measuring information as a representational and cognitive quantity, the notion of representational information and its measurement may be useful in more pragmatic settings. Some potential applications of the present theory include: (1) database analysis, (2) rule mining as in [27], (3) modeling and implementation of conceptual processes in artificial and human cognitive systems (e.g., artificial classifiers as in [10], robot perception as in [12], and AI experts), and 4) information compression.

Indeed, all four domains can significantly benefit from the kind of data compression that we believe underlies human conceptual behavior. For example, in the first five sections of this paper we have argued that RIT acts on the assumption that humans are sensitive to certain invariance patterns as the basis of concept formation and that these patterns provide the human code for optimal compression of information and parsimonious decision making. In contrast, in SIT, compressibility of information is grounded on the probability distribution function (of the random variable representing the event) as a descriptive rule that summarizes or characterizes the event space. This approach, for the reasons enumerated under Section 1, does not capture the structural or relational properties (known with certainty) about compound stimuli. However, in RIT, it is precisely these relatively few relations between the stimuli, and not the potentially complicated and only partially known statistical distributions of many stored categorization instances or categorization events (see discussion under Section 2 above), that is the basis of the compression process. Thus, RIT potentially offers more effective information compression for the above tasks.

Likewise, RIT may be useful for: (1) determining which subsets of a data set can best summarize its content (under specific size constraints), (2) to characterize how informative different types of concepts and partial concepts are to human and artificial cognitive systems, (3) to eliminate redundant information from datasets, and (4) to determine the prototypical elements of data sets when measures based on frequencies, the mean, or the median are inadequate.

For these applications, the aim is not to compute the representational information of each subset or representation of a set (hence the allusion to a "size constraint" in the previous paragraph) but of some adequate subsets or representations as dictated by the problem domain (we will illustrate this point later in this section). Indeed, computing the representational information conveyed by each representation of a "very large" well-defined category is currently intractable. That is, the problem is solvable in theory but not in practice. More specifically, no polynomial time algorithm is known for generating every subset of a finite set. Although the computation is easy, writing down all possible 2^n subsets for a set containing n objects (or for $2^n - 2$ subsets when excluding the two vacuous cases of the empty set and the set itself) will take a long time for any known computer when n is "large". The algorithmic complexity of an algorithm capable of such a feat is $O(2^n)$ or exponential [16].

To illustrate this point, it would take Tianhe-1A, the current world's fastest computer, capable of approximately 2.566 quadrillion floating point operations per second, roughly one week to compute RIT for all possible representations of a well-defined category containing 71 objects. The situation compounds when one considers that after each representation (subset) of a category is generated, its degree of invariance must be computed as well. Vigo [30] reduced the invariance detection problem to the problem of evaluating truth-tables, which is also a problem of exponential complexity in respect to the number of dimensions over which the category is defined. Although by using distributed algorithms and parallel architectures these computations would take significantly less time, there is a clear computational and practical limit to the notion of computing the quantity of representational information conveyed by every possible representation of a well-defined category containing many objects. At the same time, it should be said that for inputs of 20–30 objects and 20–30 dimensions the above computations are quite tractable and would be perceived as instantaneous when performed by a personal computer.

On the other hand, this limitation is of no consequence with respect to the kind of empirical research and the kinds of behaviors calling for explanation in cognitive science and cognitive psychology. This is due to the perceptual and memory limitations of the human cognitive system and to the highly simplified stimuli normally used in laboratory experiments. That is, in humans, fields of view are spatially constrained and the number of readily (i.e., immediately, consciously, and overtly) discriminable objects and dimensions in such spaces are usually relatively small. This makes the computation of representational information for every representation of a set of well-defined compound stimuli (i.e., a well-defined category) quite tractable. Moreover, humans are not able to discriminate more than a few dimensions consciously without considerable deliberation, unless these are overt in saliency and highly distinct. Moreover, in many cases, the most relevant representations of perceived categories are those ranging from a single object to four objects (defined over two to four dimensions) since this range presumably falls within the limit of the working memory capacity of most humans [5].

At the beginning of this Section 1 listed a number of potential applications for RIT. As mentioned, these applications are feasible when appropriate size constraints on sets of stimuli are met. Such situations arise whenever a heuristic search on a large database yields a relatively small number of alternatives that are then presented to a human agent to act upon. These alternatives (represented as well-defined categories in RIT) can then be organized, analyzed, and altered in terms of their RI. Typically, only singleton representations consisting of a single alternative (i.e., object) would be of interest in such applications (which, incidentally, would greatly simplify the computations described above). The general idea is to, after a preliminary rough search order the results in terms of their amount of representational information. This should facilitate the efficient detection and processing of information by human agents.

Systems that can benefit from the above prescription include expert systems, search engines, databases, and data mining engines: in short, any system for which it is conceivable that a query initiated by an agent may produce multiple alternatives. Next, we give an example of such a prescription which in general applies to the other potential applications listed above. Consider the task of searching for a particular webpage via an internet search engine. Let's assume that such a search produces 1000 results. Suppose that these 1000 results (as is normally the case) are laid out in a specific order by some relevance heuristic, and that they are presented 20 items (webpage links) at a time, each with a short name and description.

Furthermore, suppose that we represent each short name and description corresponding to each webpage as an object with 17 dimensions. Then, one could construct a well-defined category consisting of 17 dimensions and 20 objects. A useful way of ordering the 20 displayed webpages would be in terms of their subjective representational information content in respect to all 20 pages as a whole. Of course, this can be done for all 50 groups of 20 webpages, 20 pages at a time. Such a calculation is quite tractable for a modern home PC capable of 25 gigaflops: indeed, it would be perceptually instantaneous, as would be the generation of the truth-table of a Boolean formula in disjunctive normal form consisting of 17 variables (see the discussion on algorithmic time complexity above).

7. Conclusion and research directions

In this paper, we first introduced the notion of representational information. Representational information is the relational and qualitative information carried by subsets of a well-defined category S about S. The ability to measure representational information is useful when it is necessary to determine the essential information embedded in a finite set of objects defined in terms of dimensional values. We then introduced an original deterministic mathematical framework for the measurement and analysis of representational information based on the unification of three fundamental constructs of universal science: invariance, complexity, and information. We called this program representational information theory or RIT. To our knowledge, this marks the first time that a seamless, overt, and direct marriage between these fundamental ideas has been suggested to characterize information.

Moreover, RIT is based on a set of assumptions that are antithetical to the intuitions underlying Shannon information. First, in RIT, the mediators of information are concepts and not strings. Secondly, RIT is object driven and not event driven. The assumptions being that only an information theory grounded on concept formation can adequately account for subjective information judgments, and that only an object driven theory is capable of giving an adequate account of concept formation as the basis of information. Finally, unlike theories bound by the independence axiom of classical probability theory, RIT may account for the contextual effects that multiplicity of objects (and events) shown concurrently may have on subjective information judgments.

The deterministic nature of RIT stands also as a unique aspect of the theory, for although alternative ways of interpreting information continue to be proposed (for example see [26]), the vast majority of these continue to rely on some core notion of uncertainty. RIT assumes that the human observer knows with certainty the relational information in the stimulus set and in its representations. In contrast, since SIT is based on event probabilities, it naturally assumes that only partial information from the relevant domain is known to the observer at different points in time.

Also, it is worth noting that, although RIT is able to overcome key weaknesses in SIT, it may seem to fall short in one respect: in RIT the information conveyed by a representation R of S is about the well-defined category S. But, it has been aptly suggested that, there may be several well-defined source categories, say $S_1, \ldots, S_n$, with the same representation R. Since RIT assumes that the source category S is known along with its representation R, all these possible categories would be presumably known by the observer along with the representational signal R. Thus, R would have different information values depending on its source.

In view of this, a question arises: can RIT tell us which source set S_i an observer will act on given the representation R? We think that it can, since, at its core, RIT assumes that humans are sensitive to certain invariance patterns that account for concept formation, and that these patterns (and the way that they are detected) provide the human code for optimal information compression and for parsimonious decision making. Thus, for example, if R conveys more representational information about S_1 than it does about any other possible source set $S_1, \ldots, S_n$ known to the observer, then S_1 will be chosen. In other words, we propose that in situations involving multiple categories of origin, human receivers will follow a principle of parsimony and will focus on the category of origin for which R is most informative. Another future research direction of the current work is to design and conduct experiments to test this hypothesis.

We claimed under Section 6 that RIT may be applied to areas other than human cognition. For example, we proposed that RIT may be useful for increasing the comprehensibility and accessibility of results from database searches by organizing them according to their amount of subjective representational information. Some potential applications of this approach include: (1) database analysis; (2) modeling and implementation of conceptual processes in artificial cognitive systems (e.g., machine, robot, and AI expert, cognition) and human cognitive systems; (3) information compression, and many more. How exactly, and to what extent, these applications will benefit from RIT is still to be determined and is certainly beyond the scope of this theoretical proposal to address. We hope that the brief but specific application example outlined under Section 6 will spur an interest on the other potential applications of RIT. We believe that such efforts will help researchers gain further insight into the nature of information in their domain of interest.

Less encouraging is the fact that although the representational information measure we have proposed is effective in capturing relational information generated by context, the measure is limited in a number of ways. For one, it only applies to categories that are well-defined (i.e., dimensionally defined) or, in other words, to sets whose membership content is specified in terms of an algebraic rule. In addition, the rules considered in this paper are rules consisting of Boolean algebraic operations and binary variables representing the dimensions of the objects in the category. This lack of flexibility is compensated by the parsimony of the measure and the fact that the conceptual behavior of humans in respect to these well-defined categories have been found to be empirically lawful. Furthermore, many natural categories may be reduced to this simple binary scheme which highlights the structural sketches of categories in the complex world we live in.

Notwithstanding, a future research challenge is to extend categorical invariance theory [30] and, therefore, RIT to many-valued dimensions. Such an extension would be challenging for it would have to offer an alternative way of representing the stimulus objects in terms of rules that can capture the relationships between all the dimensional values that take part in their description. The difficulty lies in the way that negation is interpreted in Boolean algebra as a two-state operator. Some kind of multivalued logic with a different notion of negation from the classical Boolean notion may seem promising but may not be very elegant.

In spite of this limitation, we hope that the ideas introduced in this paper will pave the way to alternative information measures that take into account what we know about the ability of humans to form concepts, and about the intimate relationship (specified here) between three of the most fundamental constructs in science: invariance, complexity, and information.

Note added in proof

A generalization of categorical invariance theory which overcomes the limitations of RIT listed under sections 6 and 7, and which generalizes RIT to continuous dimensions, has been developed by the author. Also, a program that computes representational information is freely available from the author. Please, contact the author for details.

Acknowledgements

The author would like to thank Sunil Carspecken and Ragvendra Yadev for their helpful suggestions regarding this work. Correspondence and requests for materials should be addressed to Dr. Ronaldo Vigo (e-mail: vigo@ohio.edu).

References

[1] H.H. Aiken, The Staff of the Computation Laboratory at Harvard University, Synthesis of Electronic Computing and Control Circuits, Harvard University Press, Cambridge, 1951.
[2] David Applebaum, Probability and Information, Cambridge University Press, 1996.
[3] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.
[4] L.E. Bourne, Human Conceptual Behavior, Allyn and Bacon, Boston, 1966.
[5] N. Cowan, The magical number 4 in short-term memory: a reconsideration of mental storage capacity, Behavioral and Brain Sciences 24 (2001) 87–185.
[6] K. Devlin, Claude Shannon, 1916–2001, Focus: The Newsletter of the Mathematical Association of America. 21 (2001) 20–21.
[7] K. Devlin, Logic and Information, Cambridge University Press, 1991.
[8] W.K. Estes, Classification and cognition, Oxford Psychology Series, vol. 22, Oxford University Press, Oxford, 1994.
[9] J. Feldman, A catalog of Boolean concepts, Journal of Mathematical Psychology 47 (1) (2003) 98–112.
[10] D. Fisch, B. Kühbeck, B. Sick, S.J. Ovaska, So near and yet so far: new insight into properties of some well-known classifier paradigms, Information Sciences 180 (2010) 3381–3401.
[11] W.R. Garner, The Processing of Information and Structure, Wiley, New York, 1974.

[12] S. Guadarrama, A. Ruiz-Mayor, Approximate robotic mapping from sonar data by modeling perceptions with antonyms, Information Sciences 180 (2010) 4164–4188.
[13] R.V.L. Hartley, Transmission of information, Bell System Technical Journal (1928) 535–563.
[14] J.P. Hayes, Introduction to Digital Logic Design, Addison-Wesley, 1993.
[15] R.A. Higonnet, R.A. Grea, Logical Design of Electrical Circuits, McGraw-Hill, New York, 1958.
[16] E. Horowitz, S. Sahni, Computing partitions with applications to the Knapsack problem, JACM 21 (2) (1974) 277–292.
[17] J.K. Kruschke, ALCOVE: an exemplar-based connectionist model of category learning, Psychological Review 99 (1992) 22–44.
[18] W. Lenski, Information: a conceptual investigation, Information 1 (2) (2010) 74–118.
[19] R.D. Luce, Whatever happened to information theory in psychology?, Review of General Psychology 7 (2) (2003) 183–188
[20] G.L. Murphy, The Big Book of Concepts, MIT Press, 2002.
[21] R.M. Nosofsky, M.A. Gluck, T.J. Palmeri, S.C. McKinley, P.G. Glauthier, Comparing models of rule-based classification learning: a replication and extension of Shepard, Hovland, and Jenkins (1961), Memory and Cognition 22 (3) (1994) 352–369.
[22] R. Seising, Cybernetics, system(s) theory, information theory and fuzzy sets and systems in the 1950s and 1960s, Information Sciences 180 (2010) 4459–4476.
[23] C.E. Shannon, A mathematical theory of communication, System Technical Journal 27 (1948) 379–423. 623–656.
[24] C.E. Shannon, W. Weaver, The mathematical theory of communication, University of Illinois Press, Urbana, 1949.
[25] R.N. Shepard, C.L. Hovland, H.M. Jenkins, Learning and memorization of classifications, Psychological Monographs: General and Applied 75 (13) (1961) 1–42.
[26] B. Skyrms, The flow of information in signaling games, Philosophical Studies 147 (2010) 155–165.
[27] F.S. Tseng, Y. Kuo, Y. Huang, Toward boosting distributed association rule mining by data de-clustering, Information Sciences 80 (2010) 4459–4476.
[28] R. Vigo, A note on the complexity of Boolean concepts, Journal of Mathematical Psychology 50 (5) (2006) 501–510.
[29] R. Vigo, Modal similarity, Journal of Experimental and Artificial Intelligence 21 (3) (2009) 181–196.
[30] R. Vigo, Categorical invariance and structural complexity in human concept learning, Journal of Mathematical Psychology 53 (4) (2009) 203–221.
[31] R. Vigo, A dialogue on concepts, Think 9 (24) (2010) 109–120.
[32] R. Vigo, Towards a law of invariance in human conceptual behavior, in: L. Carlson, C. Hölscher, T. Shipley (Eds.), Proceedings of the 33rd Annual Conference of the Cognitive Science Society, Austin, TX: Cognitive Science Society (in press).