 Mind & Matter Vol. 4(2), pp. 169–193

Pragmatic Information in Dynamic Semantics

Peter beim Graben
Department of Clinical Language Sciences
University of Reading, United Kingdom

Abstract

In 1972, Ernst Ulrich and Christine von Weizsäcker introduced the concept of pragmatic information with three desiderata: (i) Pragmatic information should assess the impact of a message upon its receiver; (ii) Pragmatic information should vanish in the limits of complete (non-interpretable) "novelty" and complete "confirmation"; (iii) Pragmatic information should exhibit non-classical properties since novelty and confirmation behave similarly to Fourier pairs of complementary operators in quantum mechanics.

It will be shown how these three desiderata can be naturally fulfilled within the framework of Gärdenfors' dynamic semantics of Bayesian belief models. (i) The meaning of a message is its impact upon the epistemic states of a cognitive agent. A pragmatic information measure can then be quantified by the average information gain for the transition from a prior to a posterior state. (ii) Total novelty can be represented by the identical proposition, total confirmation by the logical consequence of propositions. In both cases, pragmatic information vanishes. (iii) For operators that are neither idempotent nor commuting, novelty and confirmation relative to a message sequence can be defined within Gärdenfors' theory of belief revisions. The proposed approach is consistent with measures of relevance derived from statistical decision theory and it contains Bar-Hillel's and Carnap's theory of semantic information as a special case.

1. Introduction

In his contribution to *The Mathematical Theory of Communication*, Weaver described three levels of communication "procedures by which one mind may affect another" (Shannon and Weaver 1949, pp. 95–96):

> Level A. How accurately can the symbols of communication be transmitted? (The technical problem.)
> Level B. How precisely do the transmitted symbols convey the desired meaning? (The semantic problem.)
> Level C. How effectively does the received meaning affect conduct in the desired way? (The effectiveness problem.)

By and large, these levels correspond to the semiotic dimensions discussed by Morris (1955). The technical problem with noisy communication channels can be tackled by redundant codes that introduce correlations among the elements (symbols) of the messages, i.e. by *syntax*. The semantic problem addresses correlations between the transmitted symbols and their desired meanings, i.e. *semantics*. Finally, the effectiveness problem corresponds to *pragmatics*, addressing relations between the symbols and their impact upon the users.

Although it has been argued that Shannon's theory of syntactic information were without any significance for the semantic and pragmatic dimensions (see Gernert 2006 for a discussion), Weaver claimed (Shannon and Weaver 1949, p. 98) that

> the analysis at Level A discloses that this level overlaps the other levels more than one could possible naively suspect. Thus the theory of Level A is, at least to a significant degree, also a theory of levels B and C.

It is the aim of the present study to argue in favor of this assertion. Section 2 provides a brief tutorial on classical (Shannonian) information theory and its applications. They include the approaches of Bar-Hillel and Carnap (1953, 1964) and Crutchfield (1991, 1992) for theories of semantic information and finally measures of utility and relevance that were suggested as pragmatic information measures in the framework of statistical decision theory (Polani *et al.* 2001, Weinberger 2002, van Rooij 2004, 2006).

Subsequently, Sect. 3 introduces the concept of pragmatic information proposed by E.U. and C. von Weizsäcker (1972) and E.U. von Weizsäcker (1974b) and further developed by C.F. von Weizsäcker (1974a, 1988), Gernert (1985, 1996), Kornwachs and Lucadou (1982, 1985), and Atmanspacher and Scheingraber (1990). E. U. and C. von Weizsäcker (1972) supplied three desiderata that should be obeyed by reasonable measures of pragmatic information: (i) Pragmatic information should assess the impact of a message upon its receiver; (ii) Pragmatic information should vanish in the limits of complete novelty and complete confirmation; (iii) Novelty and confirmation require a non-classical, "quantum-like" description in terms of incompatible observables. In order to meet these desiderata, dynamic semantics (Staudacher 1987, Groenendijk and Stokhof 1991, Kracht 2002, Gärdenfors 1988, 1994) will be proposed as the appropriate formal framework for describing meaning as transitions of cognitive states in Section 4.

In Sect. 5, the theory of pragmatic information will be developed in the suggested way. Section 6 presents two applications of the theory. Firstly, Bar-Hillel's and Carnap's semantic information measure will be derived. Secondly, the relation between pragmatic information in dynamic semantics and measures of relevance in statistical decision theory will be elucidated. Section 7 provides a concluding summary.

2. Information Theory

In Shannon's theory (Shannon and Weaver 1949), a discrete information source is given by a probability space $\mathcal{X}$ endowed with a probability distribution $\rho_X : \mathcal{X} \to [0,1]$, such that $p_i = \rho_X(A_i)$ is the probability of the elementary event $A_i \in \mathcal{X}$. The discrete set $\mathcal{X}$ is considered as a *repertoire* of possible messages with p_i being the probability that the message A_i is emitted. A communication system consists of the sender $\mathcal{X}$, a receiver, also regarded as an information source with repertoire $\mathcal{Y}$ and probability distribution $\rho_Y : \mathcal{Y} \to [0,1]$, and a noisy channel given by a stochastic transfer function $F : \mathcal{X} \to \mathcal{Y}$.

2.1 Syntactic Information

The aim of Shannon's theory of communication is to describe the optimal way of information transmission from the sender $\mathcal{X}$ to the receiver $\mathcal{Y}$ via the noisy channel F. Communication can be optimized by redundantly decoding the messages from $\mathcal{X}$ in such a way that a disturbed message $B_i = F(A_i)$ has the same code as the original A_i. Optimization refers to maximizing some cost function, namely the *Shannon information.*

The information content of a message A_i emitted by $\mathcal{X}$ is given by

$$I(A_i) = -\log p_i \tag{1}$$

where the base of the logarithm can be chosen arbitrarily. Usual choices are $\mathrm{ld} \equiv \log_2$ yielding information measured in bits, or $\log_M$ for a finite set of messages $\mathcal{X}$ with M elements. The latter choice normalizes the information of uniformly distributed messages ($p_i = 1/M$) to unity: $-\log_M(1/M) = \log_M M = 1$. In the sequel we shall always consider information measures normalized to the range $[0,1]$ in the latter way.

The average information of a message sent by $\mathcal{X}$ is given by its entropy

$$H(\mathcal{X}) = \sum_{A_i \in \mathcal{X}} \rho_X(A_i)\, I(A_i) = -\sum_{A_i \in \mathcal{X}} p_i \log p_i \ , \tag{2}$$

where the usual convention $p_i \log p_i = 0$ if $p_i = 0$ is applied. Similarly, one can compute the average information of the received messages as $H(\mathcal{Y})$. Since entropies are normalized by the choice of the base of the logarithm in Eq. (1), the quantity

$$G(\mathcal{X}) = 1 - H(\mathcal{X}) \tag{3}$$

measures the average certainty of a message and is called *redundancy*. $G(\mathcal{X})$ is sometimes denoted as *negentropy* (C.F. von Weizsäcker 1988).

In order to address the communication system as a whole, one considers the product space $\mathcal{X} \times \mathcal{Y}$ with the associated joint probability distribution $\rho_{XY} : \mathcal{X} \times \mathcal{Y} \to [0, 1]$. The original probabilities are regained as the marginal distributions

$$\begin{aligned} \rho_X(A_i) &= \sum_{B_j \in \mathcal{Y}} \rho_{XY}(A_i, B_j) \\ \rho_Y(B_i) &= \sum_{A_i \in \mathcal{X}} \rho_{XY}(A_i, B_j) \, . \end{aligned}$$

Equation (2) provides the joint entropy $H(\mathcal{X} \times \mathcal{Y})$ from which the *conditional entropies*

$$H(\mathcal{X}|\mathcal{Y}) = H(\mathcal{X} \times \mathcal{Y}) - H(\mathcal{Y}) \tag{4}$$

$$H(\mathcal{Y}|\mathcal{X}) = H(\mathcal{X} \times \mathcal{Y}) - H(\mathcal{X}) \tag{5}$$

are obtained. The quantity $H(\mathcal{X}|\mathcal{Y})$ measures the average uncertainty of the sender about whether or not the received message is known, thus it measures the ambiguity of the code. $H(\mathcal{X}|\mathcal{Y})$ is called *equivocation*. On the other hand, $H(\mathcal{Y}|\mathcal{X})$ measures the average uncertainty of the receiver about whether or not the emitted message is known.

The central quantity in Shannon's theory is the entropy of the sender reduced by the equivocation (or, *vice versa*, the entropy of the receiver reduced by the conditional entropy $H(\mathcal{Y}|\mathcal{X})$),

$$R(\mathcal{X}, \mathcal{Y}) = H(\mathcal{X}) - H(\mathcal{X}|\mathcal{Y}) = H(\mathcal{Y}) - H(\mathcal{Y}|\mathcal{X}) \, . \tag{6}$$

Originally, $R(\mathcal{X}, \mathcal{Y})$ was dubbed "rate of information transmission" by Shannon and Weaver (1949), but it is nowadays referred to as the *mutual information* between two arbitrary information sources $\mathcal{X}, \mathcal{Y}$ in multivariate statistics. The maximum of $R(\mathcal{X}, \mathcal{Y})$ across all possible codes defines the *channel capacity* and therefore the optimal coding for the communication.

Another important concept refers to information gain. Consider an information source $\mathcal{X}$ with an unknown probability distribution ρ. If an observer has the hypothesis that the messages from $\mathcal{X}$ are distributed according to another distribution ρ_0, he can determine the *information gain* $\Delta I(A_i)$ for the message $A_i \in \mathcal{X}$, when he becomes convinced that ρ is indeed the true distribution, as

$$\Delta I(A_i) = -\log \rho_0(A_i) - (-\log \rho(A_i)) = \log \frac{\rho(A_i)}{\rho_0(A_i)} \, . \tag{7}$$

Averaging $\Delta I(A_i)$ over all messages $A_i \in \mathcal{X}$ with respect to the true distribution ρ yields the so-called Kullback-Leibler information as the average information gain (Kullback and Leibler 1951, Kullback 1968) between the distributions ρ, ρ_0,

$$K(\rho, \rho_0) = \sum_{A_i \in \mathcal{X}} \rho(A_i) \Delta I(A_i) = \sum_{A_i \in \mathcal{X}} \rho(A_i) \log \frac{\rho(A_i)}{\rho_0(A_i)} . \tag{8}$$

Computing the Kullback-Leibler information for the product space $\mathcal{X} \times \mathcal{Y}$ under the initial hypothesis of stochastic independence, $\rho_0(A, B) = \rho_X(A)\rho_Y(B)$, yields exactly the mutual information $R(\mathcal{X}, \mathcal{Y})$.

2.2 Semantic Information

Bar-Hillel and Carnap (1953, 1964) applied Shannon's information theory to a logical calculus with a finite number of individual terms and predicates. They defined the *informativity* of a proposition by the number of logically equivalent expressions, or formulas. The informativity should be large when only a small number of equivalent formulas exists. A proposition should be less informative if there is a large number of logical equivalents.

Let A be a logical formula in Bar-Hillel's and Carnap's framework and $m(A) = \#\{P | P \leftrightarrow A\}$ the number of formulas P that are logically equivalent with A (denoted as $P \leftrightarrow A$). Assuming that these formulas are equally likely, their probability is $p(A) = 1/m(A)$. The informativity of A (see also van Rooij 2006) is then given by its Shannon information according to Equation (1):

$$I(A) = -\log p(A) = \log m(A) . \tag{9}$$

The applicability of this approach is restricted by the constraint that the number of logical formulas must be finite.

Another proposal to measure the semantic content of a message, or its meaning, was suggested by Crutchfield (1991, 1992). Consider an observer watching the output of an information source $\mathcal{X}$, namely a sequence $s = A_{i_1} A_{i_2} A_{i_3} \dots$ (cf. C.F. von Weizsäcker 1988, Chap. 5). Crutchfield assumed that the observer has a *mental model* of the ongoing process in form of a *stochastic automaton* (Crutchfield 1992, Fu 1974). Such an automaton is a tuple $\Sigma = (\mathcal{X}, Q, T, q_0)$ where $\mathcal{X}$ is the alphabet of visible messages that are to be described, Q is a set of n internal states, and $T : \mathcal{X} \to [0, 1]^{n^2}$ maps each symbol $A \in \mathcal{X}$ to an $n \times n$ stochastic matrix $T(A) = (p_{ij}(A))$ such that $p_{ij}(A)$ is the probability for the transition from state $q_j \in Q$ to state $q_i \in Q$ accepting the symbol A from the input.

Since $T(A)$ is a stochastic matrix, the columns of $T(A)$ add to unity for all messages $A \in \mathcal{X}$:

$$\sum_j p_{ij}(A) = 1\,. \tag{10}$$

Moreover, $q_0 \in Q$ is an initial state with a certain probability that characterizes complete ignorance of the observer.

Crutchfield (1992, p. 23) restricted his discussion to stochastic automata where each transition is uniquely labeled by its initial and final states q_j, q_i and the accepted message A. However, this assumption together with the normalization constraint Eq. (10) yields a deterministic automaton with $p_{ij}(A) = 1$. Therefore, Crutchfield's exposition would require another normalization condition such as $\sum_{A\in\mathcal{X}}\sum_j p_{ij}(A) = 1$ which is not appropriate for a sequence-accepting automaton but rather for a sequence-producing automaton.

Deviating from Crutchfield (1992), let us consider a stochastic automaton Σ as the mental model of an observer being in a current state $q_j \in Q$ as defined above, and receiving the message A from the information source $\mathcal{X}$. Then, the *meaning* of A in the given state q_j is the set of possible destination states $\{q_i \in Q | p_{ij}(A) > 0\}$. In this case, its semantic information is the Shannon entropy of the transition

$$\Theta_j(A) = -\sum_{q_i\in Q} p'_{ij}(A)\log p'_{ij}(A)\,, \tag{11}$$

where

$$p'_{ij}(A) = \frac{p_{ij}(A)}{\sum_{q_i\in Q} p_{ij}(A)}$$

are renormalized transition probabilities.

According to Eq. (11), the semantic information of the symbol A is zero if all $p'_{ij}(A) = 0$ for the given starting state q_j, i.e. there is no allowed transition from q_j accepting the symbol A. In this case, the mental model Σ appears to be inadequate and should therefore be revised. On the other hand, if there is only one transition from q_j to a destination state q_i, then $p_{ij}(A) = 1$ and $\Theta_j(A) = 0$ since A was anticipated by the model Σ with certainty.

In Crutchfield's semantics of automata, messages act as operators upon the states of an automaton and the meaning of a symbol is the result of such an operation. Additionally, meaning and semantic information are contextually defined with respect to both the valid mental model Σ and the current state q_j of the observer (cf. Staudacher 1987, Atmanspacher *et al.* 1992). We shall see in Sect. 4 that these assumptions actually form the core of dynamic semantics.

2.3 Pragmatic Information

Any measure of pragmatic information content must be contextually defined (Atmanspacher *et al.* 1992) with respect to a given agent and to her beliefs, desires and goals at a given instance of time. This is similar to the discussion of Crutchfield's theory of semantic information where an observer is in a particular state q_j of a particular mental model Σ. The appropriate framework of such an account is *statistical decision theory*.[1]

Statistical decision theory describes the possible decisions and actions of an agent in a particular *epistemic state*. Let X be the agent's state space, $x \in X$ an epistemic state with probability $\rho(x)$ and Y the set of decisions (van Rooij 2004). Decisions are evaluated according to a *utility function* $u : X \times Y \to \mathbb{R}$, such that $u(x, y)$ is the utility of the decision y in state x. An optimal decision $y^* \in Y$ maximizes the utility: $u(x, y^*) \geq u(x, y)$ for all $y \in Y$.

One important problem in statistical decision theory is to find an appropriate utility function for rationally and consistently (or honestly) behaving agents. This problem was solved by Bernardo (1979) and by von Weizsäcker (1988), together with Drieschner, using information theory. They proved that the best utility function is simply a linear function of the information content (Eq. (1)) of the agent's epistemic states. The expected utility for an ensemble of alternative decisions hence equals Shannon's entropy in Eq. (2).

This result was applied in linguistics by van Rooij (2004) (see also van Rooij 2006) in order to measure the *relevance* of questions. Van Rooij (2004) found that the *utility value* of a particular answer q' to a question Q' is given by the Kullback-Leibler distance (Eq. (8)) between the *a priori* distribution of the answers $\rho(q)$ to another question Q and the distribution conditionalized with respect to the given answer q', $\rho(q|q')$. Then, the mean utility value of the question Q' is obtained by its mutual information with respect to Q,

$$R(Q, Q') = \sum_{q \in Q} \sum_{q' \in Q'} \rho(q \wedge q') \log \frac{\rho(q \wedge q')}{\rho(q)\rho(q')} ,$$

where $q \wedge q'$ denotes the logical conjunction of q and q' (see Sect. 4).

A similar result was obtained by Polani *et al.* (2001) who quantified the relevance of an agent's epistemic states $x \in X$ for the optimal decision problem by the mutual information (Eq. (6)) between the distribution of states $\rho(x)$ and a uniform distribution of optimal actions $y \in Y^*(x)$ given by

[1] A different approach by Frank (2003) rests on algorithmic complexity. Two messages are called *pragmatically equivalent* if they conduct the same behavior of their recipient. The pragmatic information of a message is then the length of the shortest message in the class of all pragmatically equivalent messages.

$$\rho(y|x) = \begin{cases} 1/\#(Y^*(x)) & : \quad y \in Y^*(x)\,, \\ 0 & : \quad \text{otherwise}\ . \end{cases}$$

Finally, the approach of Weinberger (2002) provides a measure of pragmatic information of a message ensemble by the mutual information between the messages and the decisions of the agent after she has received them.

3. Three Desiderata for Pragmatic Information

E.U. and C. von Weizsäcker (1972) formulated three desiderata that should be fulfilled by any reasonable measure of pragmatic information; see also E.U. von Weizsäcker (1974). Let us briefly discuss these requirements and how they were resumed by other authors in the following subsections (for a comprehensive review see Gernert 2006).

3.1 Impact upon a Recipient

First, pragmatic information should assess the impact of a message upon its receiver (E.U. and C. von Weizsäcker 1972, p.541, translation by the author):

> Pieces of information . . . are intended to act. By definition they act upon their receivers and change them informationally. *In particular, after the arrival of a message the receiver's expectation probability for a related message will usually not be the same as before.*

This idea was taken up by Kornwachs and Lucadou (1985, p. 86) who stated that

> ... the action, provoked by information in a system is not only a simple reaction – it can alter the receiving system as a whole without direct reactions: thus it can alter the potential dispositions of a system.

Later, Gernert (1996, p. 150) specified this claim saying that "pragmatic information is characterized by the property to alter the structure and/or the behavior of the receiving system". Following Gernert (1996, 2006), structural or behavioral changes of the receiver can be described by *graph grammars*.

3.2 Novelty and Confirmation

The second desideratum states that pragmatic information consists of two components, novelty and confirmation. E.U. and C. von Weizsäcker (2006) write:

> We proposed that meaningful information consists of *two* mutually complementary components, *novelty* and *confirmation* ... *novelty* relates to entropy, *confirmation* to negentropy.

Identifying novelty with Shannon entropy H as a measure of randomness and confirmation with negentropy or *redundancy*, $G = 1 - H$ (Eq. (3)), pragmatic information should behave similar to a measure of complexity that depends non-monotonically on randomness and is globally concave (Atmanspacher *et al.* 1992). The relationship between pragmatic information and complexity was also addressed by E.U. and C. von Weizsäecker (1972) and E.U. von Weizsäcker (1974). Gernert (1996, 2006) suggested to consider novelty and confirmation as independent variables.

An early, though rather problematic, attempt to measure novelty N and confirmation C is due to Kornwachs and Lucadou (1982). They formulated for the first time the so-called "product formula" for pragmatic information:

$$S = N \cdot C \,, \tag{12}$$

resembling an ostensible measure of complexity suggested by Shiner *et al.* (1999) that was criticized by Crutchfield *et al.* (2000).

Considerable progress was made by Gernert (1996, 2006) who suggested to measure novelty as the dissimilarity between a message and its recipient's actual knowledge. By contrast, confirmation has to be measured by the dissimilarity between the message and the agent's expectations or goals. Similarity can be measured by a metric defined on sets of graphs generated by a graph grammar. Expressing the agent's knowledge in terms of statistical decision theory as an epistemic state, one can employ Shannon's information theory for the evaluation of similarity. This issue will be reconsidered in Sect. 5 in the framework of dynamic semantics.

Another way for measuring the pragmatic information content of a message by a resulting change of efficiency of the receiver's behavior was originally proposed by Gernert (1985) and subsequently applied to a physical system by Atmanspacher and Scheingraber (1990). Let η_{prior} be the efficiency of the receiver before and η_{post} after the reception of a message. The pragmatic information can then be assessed by

$$S = \frac{\eta_{\text{post}} - \eta_{\text{prior}}}{\eta_{\text{post}}} \,. \tag{13}$$

This definition can be related to the "computational mechanics" of stochastic automata. Crutchfield (1992) defined the efficiency of a model description Σ as

$$\eta = \frac{||\Sigma|| + ||E||}{||s||} \,. \tag{14}$$

Here, $||s||$ is the length of the optimal symbolic encoding s of the pattern to be modeled, $||\Sigma||$ is the length (the algorithmic complexity) of a description of the stochastic automaton Σ reproducing s, and $||E||$ is the length of the description of the residual error not accounted for by the model Σ. From the point of view of pragmatic information, $||\Sigma||$ assesses confirmation while $||E||$ measures novelty.

3.3 Non-Classicality

A "formal speculation" led E.U. and C. von Weizsäcker (1972) and E.U. von Weizsäcker (1974) to the idea that novelty and confirmation could be considered as Fourier pairs in the sense of *complementary* observables requiring a non-classical, quantum-like treatment. Their argument was as follows. If total novelty is represented by a delta function signal at some time t_0, its Fourier transform is a uniform distribution over all frequencies. In the same picture, complete confirmation would be a constant signal for all times, and its Fourier transform a delta peak at a certain frequency f_0. Therefore, total novelty and complete confirmation are a Fourier pair. In quantum theory, complementary observables are such Fourier pairs, obeying Heisenberg's uncertainty relations. (However, the converse is usually not the case: Fourier pairs are not necessarily complementary in the sense of quantum theory.)

This speculation was further strained by Kornwachs and Lucadou (1982, 1985) in order to justify Eq. (12). If novelty and confirmation were complementary observables they should fulfill an uncertainty relation, where the pragmatic information S in Eq. (12) provides the lower bound of the product of N and C. A non-classical description of pragmatic information for cognitive operations was later suggested by Gernert (2000) and Atmanspacher and Filk (2006). In Sect. 5.3 we shall come back to this point.

4. Dynamic Semantics

Statistical decision theory describes the behavior of a cognitive agent in a particular epistemic state that can change under the influence of received messages. This resembles the description of an observer watching an information source and building computationally efficient (mental) models from those observations. Both Crutchfield (1992) and Gernert (2000) proposed that the meaning of a message is described by an operation on the epistemic state space of the agent.

In this Section the approach of *dynamic semantics* will be suggested as a unifying account for these proposals. According to Groenendijk and Stokhof (1991),

> the meaning of a sentence does not lie in its truth conditions, but rather in the way it changes (the representation of) the information of the interpreter. The utterance of a sentence brings us from a certain state of information to another one.

Similarly, Kracht (2002, p. 217) states that

> dynamic semantics is called 'dynamic' because it assumes that the meaning of a sentence is not its truth condition but rather its impact on the hearer.

In the following, let us briefly review Gärdenfors' theory of *belief models* (Gärdenfors 1988, 1994) as a formal framework for a theory of pragmatic information that fulfills the three desiderata of E.U. and C. von Weizsäcker (1972). Mathematically, Gärdenfors' theory is an application of category theory and can be easily recast in terms of a generalized quantum theory (Atmanspacher *et al.* 2002, Römer 2004).

4.1. Belief Models

The generalized framework of Atmanspacher *et al.* (2002) considers a set X as a state space and functions from X to X in $\mathrm{Mor}(X)$. Particular functions from a countable, discrete subset $\mathcal{A} \subseteq \mathrm{Mor}(X)$ are called *observables.* Observables can be concatenated, i.e. iteratively invoked, such that $(A \circ B)(x) = A(B(x)) = A(y)$ if $y = B(x)$ for all $x \in X$. The product $AB = A \circ B$ of observables is associative: $A(BC) = (AB)C$. Atmanspacher *et al.* (2002) supplied a number of axioms for the properties of observables. The set $\mathcal{A}$ becomes a monoid if it is closed with respect to the concatenation product and contains a neutral element $\top$ such that

$$\top \circ A = A \circ \top = A \tag{15}$$

for all $A \in \mathcal{A}$. A subset $\mathcal{P} \subset \mathcal{A}$ is a commutative sub-monoid if all elements in $\mathcal{P}$ commute with each other:

$$\mathcal{P} = \{A, B \in \mathcal{A} | AB = BA\} . \tag{16}$$

Projectors are idempotent observables $A \in \mathcal{P}$ obeying

$$A^2 = AA = A . \tag{17}$$

In Gärdenfors' dynamic semantics, the set X is regarded as the *space of epistemic states* of a particular cognitive agent. Elements $x, y, z \in X$ are called *epistemic states*, or simply *belief states.* In the sequel, we shall refer to observables $U, V \in \mathcal{A} \subseteq \mathrm{Mor}(X)$ as to *epistemic operators.* Commutative projectors $A, B \in \mathcal{P} \subseteq \mathcal{A}$ are called *propositions.* Their product

$$A \wedge B = AB = BA = B \wedge A \tag{18}$$

is called the *conjunction* of A and B.

An important notion in Gärdenfors' theory is that of *acceptance*. A proposition $A \in \mathcal{P}$ is said to be *accepted in state* $x \in X$ (or, likewise, x accepts A) if

$$A(x) = x\,. \tag{19}$$

This means that the state x is a fixed point of A. Since all propositions are idempotent, the following holds. Let $y = A(x)$ for some $x \in X$, then $A(y) = A(A(x)) = A^2(x) = A(x) = y$, i.e. after applying A to any state x the result $y = A(x)$ accepts A. Thus, Eq. (19) has the straightforward interpretation that an agent, who does not believe that A is true in state x, does believe so after being informed by the proposition A.

Another important notion, *logical consequence*, is defined as follows: A proposition B is a logical consequence of a proposition A if

$$B \wedge A = A \wedge B = A\,. \tag{20}$$

In this case, $y = A(x)$ entails $B(y) = B(A(x)) = A(x) = y$, such that B is accepted whenever A is accepted in an epistemic state (but not *vice versa*).

Gärdenfors introduced other axioms defining logical connectives such as negation ($\neg A$) or disjunction ($A \vee B$), which give rise to an interpretation of the theory in terms of propositional logic. A map $\mathbf{I}$, called *interpretation function*, assigns to each logical formula φ a proposition $\mathbf{I}_\varphi = A \in \mathcal{P}$ such that $\mathbf{I}$ is a homomorphism with respect to the connectives. Since a proposition $A = \mathbf{I}_\varphi$ is an operator $A : X \to X$, we immediately obtain that the *meaning* of a sentence (a formula of classical propositional logic) is its associated epistemic operator acting on the belief states of a cognitive agent.

A set of commutative projectors $\mathcal{P}$ acting on an agent's epistemic state space X that obeys Gärdenfors' axioms for the suggested logical interpretation constitutes a *belief model* $(X, \mathcal{P})$ (Gärdenfors 1988, 1994).

4.1. Bayesian Belief Models

So far we have considered a deterministic setup where a proposition $A \in \mathcal{P}$ is either accepted or not accepted with certainty in an epistemic state $x \in X$. In case of acceptance, we have the fixed point equation (19), $A(x) = x$. This can be equivalently expressed by assigning a probability distribution of the propositions to state x such that

$$\rho_x(P) = \begin{cases} 0 & : \quad P(x) \neq x \\ 1 & : \quad P(x) = x \end{cases} \tag{21}$$

is the probability that a proposition P is accepted in state x. This deterministic description is generalized by allowing for arbitrary distributions $\rho(P) \in [0, 1]$, where $\rho(P) = 1$ if P is accepted with certainty by the distribution ρ and $\rho(P) = 0$ if $\neg P$ is accepted with certainty by ρ. In all other cases, $0 < \rho(P) < 1$ means that P is believed to be true with probability $\rho(P)$. Thus, the state x is replaced by a map $\rho : \mathcal{P} \to [0,1]$ and the state space X is replaced by the set of all probability distributions $\mathfrak{S} = \{\rho | \rho : \mathcal{P} \to [0,1], \sum_{P \in \mathcal{P}} \rho(P) = 1\}$.[2]

In Sect. 4.1 we have modeled propositions as operators mapping epistemic states to other epistemic states. In the probabilistic (Bayesian) description, states are probability distributions. Again, propositions can be interpreted as operators (Gärdenfors 1988, Chap. 5): Let $A \in \mathcal{P}$ be a proposition and $\rho : \mathcal{P} \to [0,1]$ be a probability distribution over $\mathcal{P}$ such that $\rho(A) > 0$ and $\sum_{P \in \mathcal{P}} \rho(P) = 1$. We define the image of ρ under the impact of A, $A\rho =: \rho_A$, by the *conditionalization*

$$\rho_A(P) = \frac{\rho(P \wedge A)}{\rho(A)} =: \rho(P|A) \tag{22}$$

for all $P \in \mathcal{P}$. In the transformed state ρ_A, A is accepted to be true: $\rho_A(A) = \rho(A|A) = 1$.

A Bayesian probabilistic belief model is a pair $(\mathfrak{S}, \mathcal{P})$ where (i) $\mathcal{P}$ is a set of propositions from a belief model $(X, \mathcal{P})$ and propositions $A \in \mathcal{P}$ act on statistical states according to Eq. (22), and (ii) $\mathfrak{S}$ is a set of probability distributions such that X is identified with the pure states in $\mathfrak{S}$ (Gärdenfors 1988, Williams 1980).

After conditionalization with respect to A, only the proposition A and its logical consequences (i.e. all $P \in \mathcal{A}$ with $PA = A$) are accepted in the state ρ_A. As $\rho_A(B) = 0$ for $AB \neq A$, the denominator of Eq. (22) vanishes for conditionalizing with respect to new evidence for B. Therefore, a model for belief revision is required.

4.3. Revision-Extended Belief Models

Belief models are monotonic, i.e. propositions which are already accepted remain accepted during the dynamics of epistemic states. This is a consequence of the commutativity of propositions and can be easily checked: Let A be accepted in state x (i.e. $A(x) = x$) and let $B(x) = y$, such that B is learned during the transition $x \to y$. Then $A(y) = A(B(x)) = (A \wedge B)(x) = (B \wedge A)(x) = B(A(x)) = B(x) = y$, which means that A and B are both accepted in the new state y.

Such an account is not appropriate when mental models have to be revised either by evidence against convictions or to improve their efficiency (Crutchfield 1991, Gernert 2000). While propositions are commutative operators, this is generally not the case for belief revision processes.

[2] This construction is well-known in algebraic quantum theory, where the space of statistical states contains the positive, normalized, and linear expectation value functionals. Applying such a functional to a projector P yields the likelihood to observe the eigenvalue “true” for P. The *pure* states in $\mathfrak{S}$ obey Eq. (21).

Gärdenfors (1988) models belief revisions by means of a further set of epistemic operators $\mathcal{R} \subset \mathcal{A}$. He defines a revision-extended belief model as a triple $(X, \mathcal{P}, \mathcal{R})$ where (i) $(X, \mathcal{P})$ is a belief model, (ii) $\mathcal{P} \subseteq \mathcal{R}$, and (iii) for each $A \in \mathcal{P}$ there is an $A^* \in \mathcal{R} \setminus \mathcal{P}$ which describes the revision dynamics by further axioms.

In order to illustrate this dynamics, let us consider an agent in a belief state x that accepts proposition $A =:$ "the moon consists of Gorgonzola". Its revision is $A^* =:$ "the moon does not consist of Gorgonzola". Another proposition might be $B =:$ "the moon is a big rock". Since x accepts A, the application of B, $B(x)$, leads to the *absurd state* $o \in X$ where all propositions are accepted (Gärdenfors 1988), including $BA =:$ "the moon is a big rock consisting of Gorgonzola". This state is invariant under the revision A^*, hence $(A^*B)(x) = o$. On the other hand, the product BA^* applied to x yields $B(y)$ where $y = A^*(x)$ accepts the revision of A. Therefore, $BA^*(x) \neq o$ because BA^*, "the moon does not consist of Gorgonzola, it is rather a big rock", can be consistently accepted. Thus, $A^*B \neq BA^*$, i.e. belief revisions and propositions do not commute – they are incompatible or even *complementary.*

Revision-extended belief models can also be defined for the Bayesian account by introducing a *revision function* $* : \mathcal{P} \to \mathcal{R}$ which assigns to a proposition $A \in \mathcal{P}$ its revision $A^* \in \mathcal{R}$. However, the revision A^* of A can only act on a state ρ according to the conditionalization rule Eq. (22) if $\rho(A) > 0$, i.e. if A is consistent with the belief expressed in ρ. If A is not consistent with ρ, A^* must act on ρ differently. This difference is captured by the following concepts (Gärdenfors 1988, Chap. 5).

A set $\mathfrak{O}(\rho)$ associated to a statistical state $\rho \in \mathfrak{S}$ is called an *ordinal family* for ρ, if there is a well-ordering $\prec$ on $\mathfrak{O}(\rho)$ such that (i) $\rho = \min \mathfrak{O}(\rho)$ and $\rho_\perp = \max \mathfrak{O}(\rho)$, and (ii) for all $A \in \mathcal{P}$ there is some $\rho^a \in \mathfrak{O}(\rho)$ such that $\rho^a(A) > 0$ and $\rho^a \neq \rho_\perp$, if not $\neg A = \top$. Here $\perp$ denotes the contradiction and the neutral element $\top$ is interpreted as a tautology of the belief model.

Given such an ordinal family $\mathfrak{O}(\rho)$, the revision A^* acts on ρ as

$$\rho_{A^*} := \rho^*_A \tag{23}$$

where ρ^* is the *first* element ρ^a of $\mathfrak{O}(\rho)$. Thereby, revision is reduced to conditionalization (although a new probability distribution ρ^* is used instead of ρ with $\rho^*(A) > 0$). It is clear that $\rho^* = \rho = \min \mathfrak{O}(\rho)$ for $\rho(A) > 0$, which means that the effect of a revision A^* equals the effect of a proposition A if A is consistent with the epistemic state ρ. If more than one well-ordering $\prec$ is possible on $\mathfrak{O}(\rho)$, the revision function defined by Eq. (23) is not uniquely determined. There exist many different revision functions for one belief model. A proper choice of one of them depends on contexts such as the *epistemic entrenchment* of propositions (Gärdenfors

1988, Chap. 4; see also Goodman 1983), the agent's preferences and her readiness to give up a belief, her goals in certain circumstances, and other pragmatic conditions.

5. Theory of Pragmatic Information

It will now be demonstrated how the three desiderata for measures of pragmatic information (E.U. and C. von Weizsäcker 1972, E.U. von Weizsäcker 1974) can be met in the framework of dynamic semantics. For this purpose we concentrate on behavioral changes of an agent and disregard structural changes.

5.1 Impact upon a Recipient

First of all, the meaning of a message is its impact upon the belief states of a cognitive agent: Messages act as operators on the agent's epistemic state space. For the Bayesian account, where belief states are given by probability distributions over propositions, a received proposition A transforms an *a priori* distribution ρ into an *a posteriori* distribution ρ_A according to the conditionalization rule Eq. (22). The impact of that message can therefore be assessed by its average information gain, i.e. by the Kullback-Leibler information Eq. (8),

$$K(\rho_A, \rho) = \sum_{P \in \mathcal{P}} \rho_A(P) \log \frac{\rho_A(P)}{\rho(P)} \,. \tag{24}$$

However, one has to observe that the denominator of Eq. 22 does not vanish. Therefore, it is plausible to suggest

$$S_\rho(A) = \begin{cases} K(\rho_A, \rho) & : \quad \rho(A) > 0 \\ 0 & : \quad \rho(A) = 0 \end{cases} \tag{25}$$

as a measure of the pragmatic information of a proposition A in the belief state ρ of a particular agent. This definition is obviously contextual as required by Gernert (1996, 2006).

5.2 Novelty and Confirmation

In order to measure novelty and confirmation, Gernert (1996, 2006) proposed to determine the similarity between the message and the agent's epistemic state or her goals and desires, respectively. Dynamic semantics supplies a direct way for characterizing at least total novelty and complete confirmation. A proposition A will be called *totally novel* if there is no belief state ρ that can be conditionalized by A. In other words, A acts on all states as the tautology $A = \top$, yielding for all $\rho \in \mathfrak{S}$:

$$A\rho = \rho \,. \tag{26}$$

Inserting Eq. (26) into Eqs. (24) and (25), one obtains $S_\rho(A) = 0$, i.e. the pragmatic information of a totally novel message, that cannot be understood by the agent, vanishes. On the other hand, complete confirmation can easily be defined by the concept of logical consequence in Eq. (20). A proposition B will be called *completely confirmed* if B is the logical consequence of another proposition A. In that case, we have

$$(\rho_A)_B(P) = \frac{\rho_A(PB)}{\rho_A(B)} = \frac{\rho(PBA)}{\rho(A)} \cdot \frac{\rho(A)}{\rho(BA)} = \frac{\rho(PBA)}{\rho(BA)} = \rho_{BA}(P) \, ,$$

and

$$\rho_{BA}(P) = \frac{\rho(P(BA))}{\rho(BA)} = \frac{\rho(PA)}{\rho(A)} = \rho_A(P) \, ,$$

since $BA = A$, and hence $S_{\rho_A}(B) = 0$. As a result, the pragmatic information of a completely confirming message vanishes as well.

The requirement that pragmatic information should be non-negative and globally concave as a function of randomness is met by the properties of the Kullback-Leibler information (Kullback 1968, Chap. 2.3).

5.3 Non-Classicality

Since total novelty or complete confirmation are too restrictive for useful applications of the theory,[3]. one has to look for notions of novelty and confirmation relative to a given information source, or to a string of concatenated messages (E.U. and C. von Weizsäcker 1972, E.U. von Weizsäcker 1974, C.F. von Weizsäcker 1988, Kornwachs and Lucadou 1982).

Let $s = \prod_k A_k$ be a finite or infinite sequence of propositions, called a *text*. Consider for instance the sequence $s = ABBA$ of propositions $A, B \in \mathcal{P}$ from a Bayesian belief model. Texts have to be read from right to left (the Arabian way) because the rightmost proposition acts first on a current epistemic state. Since propositions are commuting projectors, $s = ABBA = AABB$ is the same text as before. Therefore, novelty cannot be defined relative to a text s of propositions, because there is no "first" proposition in s. After reordering s to $s = AABB$, the idempotence property of propositions yields $s = AABB = A^2B^2 = AB$. Hence, there is also no relative notion of confirmation with respect to s.

In order to define relative novelty and relative confirmation we, thus, have to consider epistemic operators that are not propositions. One example is provided by the belief revisions encountered in Sects. 4.2 and 4.3. For each proposition $A \in \mathcal{P}$ there is a revision $A^* \in \mathcal{R}$ usually not commuting with other propositions.

[3] E.U. and C. von Weizsäcker (1972, p. 544) refer to an agent who does not understand at all as a "hermeneutic monster".

Another important class of examples are *anaphers* as discussed by Staudacher (1987), Groenendijk and Stokhof (1991), and Kracht (2002). Consider the three propositions $A =:$ "John sat at the table", $B =:$ "George came in", and $C =:$ "he was wearing a hat". Then the pronoun "he" has conflicting interpretations for the conjunctions ABC and BAC, respectively (Frisch *et al.* 2004). Anaphers of this kind are described, either by *predicate logic with anapher* (Staudacher 1987) or by *dynamic predicate logic* (Groenendijk and Stokhof 1991, Kracht 2002), as non-commutative contextual operators.

Therefore we have to extend our theory of pragmatic information to arbitrary epistemic operators $U, V \in \mathcal{A}$ with $\mathcal{P} \subset \mathcal{R} \subset \mathcal{A}$ such that $U^2 \neq U$ and $UV \neq VU$. How do such operators act on epistemic states? For propositions we have the conditionalization rule Eq. (22), which applies also to belief revisions since the revision function $* : A \mapsto A^*$ is defined by a "quantum leap" in the state space, where $\rho_{A^*} = \rho^*_A$ and ρ^* is the smallest member of the contextually given ordinal family $\mathfrak{O}(\rho)$ with $\rho^*(A) > 0$. To be as conservative as possible, we assume that any epistemic operator V with $\rho(V) > 0$ can conditionalize a belief state by

$$\rho_V(U) = \frac{\rho(UV)}{\rho(V)} =: \rho(U|V) \tag{27}$$

if the normalization constraint

$$\sum_{U \in \mathcal{A}} \rho(UV) = \rho(V) \tag{28}$$

holds. Note that the ordering in the nominator of Eq. (27) must be carefully observed.

Now it is convenient to introduce two quantities that are formal analogues to the variance and the covariance. We define the *operator variance*

$$\mathrm{vax}_\rho(U) = \rho(U^2) - \rho(U)^2 \tag{29}$$

and the *conditional operator covariance*

$$\mathrm{cox}_\rho(U|V) = \rho(UV) - \rho(U)\rho(V)\,. \tag{30}$$

both for a given belief state ρ and for $U, V \in \mathcal{A}$. Note that ρ is a probability distribution over epistemic operators and not an expectation value functional, and that epistemic operators belong to an abstract monoid and not to an algebra. Therefore, Eqs. (29) and (30) define neither a variance nor a covariance in the sense of linear statistics.

Because epistemic operators are not linear maps, the operator variance and the conditional operator covariance can assume arbitrary signs. Only for propositions $A, B \in \mathcal{P}$ we have $\mathrm{vax}_\rho(A) = 0$ due to idempotence and

$\text{cox}_\rho(A|B) = \text{cox}_\rho(B|A)$ due to commutativity. Moreover, if proposition B is the logical consequence of proposition A, we obtain $\text{cox}_\rho(B|A) = \rho(A) \cdot \rho(\neg B)$.

What is the meaning of these general operators ? Let us first consider an operator with

$$\rho_U(U) > \rho(U)\,. \tag{31}$$

Such an operator supports itself by self-confirmation. It can be called *conviction* because the agent's probability of believing U increases under repetition of U. Computing the operator variance of U using Eq. (31) yields

$$\text{vax}_\rho(U) > 0\,. \tag{32}$$

As the counterpart to Eq. (31), consider now

$$\rho_U(U) < \rho(U)\,. \tag{33}$$

An agent receiving U over and over again becomes more and more suspicious about its credibility. One may call U *propaganda* in this case. From Eq. (33) it follows that

$$\text{vax}_\rho(U) < 0\,. \tag{34}$$

Now we are finally able to define relative novelty and confirmation. An operator $U \in \mathcal{A}$ is called novel relative to an operator $V \in \mathcal{A}$ and an epistemic state ρ if

$$\text{cox}_\rho(U|V) = 0\,. \tag{35}$$

Correspondingly, an operator $U \in \mathcal{A}$ is called relatively confirmed by an operator $V \in \mathcal{A}$ in an epistemic state ρ if

$$\text{cox}_\rho(U|V) > 0\,. \tag{36}$$

Additionally, the notion of relative *disconfirmation* is introduced by

$$\text{cox}_\rho(U|V) < 0\,. \tag{37}$$

In this case, $U \in \mathcal{A}$ will be considered disconfirmed by $V \in \mathcal{A}$ in the state ρ. These definitions are justified by

$$\text{cox}_\rho(U|V) = 0 \iff \rho_V(U) = \rho(U)$$

for novelty, and

$$\text{cox}_\rho(U|V) \lessgtr 0 \iff \rho_V(U) \lessgtr \rho(U)$$

for confirmation or disconfirmation, respectively.

Relative novelty as just defined corresponds closely to the total novelty defined in Sect. 5.2 because it does not alter the epistemic state upon conditionalization. Therefore, it provides vanishing pragmatic information $S_\rho(V) = 0$ where the sum in Eq. (24) extends over all epistemic operators

$$K(\rho_V, \rho) = \sum_{U \in \mathcal{A}} \rho_V(U) \log \frac{\rho_V(U)}{\rho(U)} \,. \tag{38}$$

The sum in Eq. (38) can be decomposed into three contributions for novelty $N_\rho(V)$, confirmation $C_\rho(V)$, and disconfirmation, $D_\rho(V)$,

$$\begin{aligned}
N_\rho(V) &= \sum_{U:\mathrm{cox}(U|V)=0} \rho_V(U) \log \frac{\rho_V(U)}{\rho(U)} \\
C_\rho(V) &= \sum_{U:\mathrm{cox}(U|V)>0} \rho_V(U) \log \frac{\rho_V(U)}{\rho(U)} \\
D_\rho(V) &= \sum_{U:\mathrm{cox}(U|V)<0} \rho_V(U) \log \frac{\rho_V(U)}{\rho(U)}
\end{aligned}$$

for $\rho(V) > 0$, such that $S_\rho(V) = N_\rho(V) + C_\rho(V) + D_\rho(V)$. Obviously, $N_\rho(V) = 0$ since $\rho_V(U) = \rho(U)$ for all $U \in \mathcal{A}$. This is at variance with the product formula Eq. (12) of Kornwachs and Lucadou (1982) which, in our terminology, would read $S_\rho(V) = N_\rho(V) \cdot C_\rho(V)$.[4]

6. Applications

The theory of pragmatic information introduced above has two obvious applications. First, it allows us to derive Bar-Hillel's and Carnap's theory of semantic information as a special case (Bar-Hillel and Carnap 1953, 1964). Second, one can demonstrate its consistency with statistical decision theory and its suggested measures of relevance (Polani *et al.* 2001, Weinberger 2002, van Rooij 2004, 2006).

6.1 Semantic Information

In order to derive Bar-Hillel's and Carnap's concept of semantic information, consider a Boolean algebra $\mathcal{P}$ consisting of only two epistemic operators, $\mathcal{P} = \{\top, \bot\}$, where $\top$ denotes the tautology and $\bot$ the contradiction. Let

$$\rho(\top) = \beta; \quad \rho(\bot) = 1 - \beta \tag{39}$$

[4]The role of disconfirmation was not discussed by Kornwachs and Lucadou (1982).

be an *a priori* probability distribution over $\mathcal{P}$. Conditionalization with $\top$ yields the *a posteriori* state

$$\rho_\top(\top) = 1; \quad \rho_\top(\bot) = 0\,. \tag{40}$$

Using Eqs. (24) and (25), the pragmatic information of $\top$ is

$$S_\rho(\top) = -\log\beta = I(\top) \tag{41}$$

which expresses the same situation as Eq. (9).

6.2 Measures of Relevance

Basically, the suggested measure of pragmatic information for a message is the expected information gain associated with the transition of an agent's Bayesian belief states induced by the reception of the message. Similar measures were suggested by van Rooij (2004) and Williams (1980). As measures of relevance, Polani *et al.* (2001), Weinberger (2002), and van Rooij (2004) suggested the mutual information between all random events that are relevant for the epistemic state of an agent.

The expectation value of $S_\rho(V)$ over all operators is given by

$$S_\rho = \sum_{V\in\mathcal{A}} \rho(V) S_\rho(V) = \sum_{U,V\in\mathcal{A}} \rho(UV) \log\frac{\rho(UV)}{\rho(U)\rho(V)} \tag{42}$$

which is simply the mutual information between two operators in $\mathcal{A}$.

7. Summary

This study resumes a theory of pragmatic information originally proposed by E.U. and C. von Weizsäcker (1972) and E.U. von Weizsäcker (1974) and further developed by C.F. von Weizsäcker (1974, 1988), Gernert (1985, 1996), Kornwachs and Lucadou (1982, 1985) and Atmanspacher and Scheingraber (1990). E.U. and C. von Weizsäcker (1972) formulated three desiderata to be met by a reasonable measure of pragmatic information. (i) Pragmatic information should assess the impact of a message upon its receiver. (ii) In the limits of non-interpretable total novelty and complete confirmation, the pragmatic information should vanish. (iii) Novelty and confirmation behave as Fourier pairs of complementary operators, so pragmatic information should exhibit non-classical properties.

The first desideratum was previously addressed by approaches from computational mechanics (Crutchfield 1991, 1992) and statistical decision theory (Polani *et al.* 2001, Weinberger 2002, van Rooij 2004). Within the unifying framework of dynamic semantics (Staudacher 1987, Groenendijk and Stokhof 1991, Kracht 2002, Gärdenfors 1988, 1994), the meaning of a message is its impact upon the space of epistemic states of a cognitive agent (a similar demand was required by Gernert (2000)). Thus, pragmatic information is a contextual notion. It refers to a particular agent in a particular belief state (Gernert 1996, 2006, Atmanspacher *et al.* 1992, 1997, Atmanspacher and Wiedenmann 1999).

The dependence of pragmatic information on novelty and confirmation is related to various measures of complexity (Grassberger 1986, Crutchfield and Young 1989, Wackerbauer *et al.* 1994, Badii and Politi 1997). Pragmatic information vanishes for total novelty (randomness) and for complete confirmation (regularity) and is globally concave for intermediate values of randomness. If the pragmatic information value of a message is defined by its Kullback-Leibler information, these properties do naturally hold. Moreover, since the average pragmatic information is given by the mutual information between messages, this measure is closely related to measures of complexity suggested by Saparin *et al.* (1994), Quian Quiroga *et al.* (2000, 2002), Badii *et al.* (1991), and Shalizi *et al.* (2004).

A classical description of epistemic operators in dynamic semantics prevents defining novelty and confirmation relative to a text or an information source. Therefore, belief revisions, convictions, propaganda, anaphers and other cognitive operators (Goodman 1983, Gärdenfors 1988, 1994, Gernert 2000, Staudacher 1987, Groenendijk and Stokhof 1991, Kracht 2002, and Atmanspacher and Filk 2006) must be taken into account for a proper treatment of pragmatic information. These operators are generally non-commutative and, therefore, incompatible or even complementary.

The proposed theory is consistent with measures of relevance derived from statistical decision theory (Polani *et al.* 2001, Weinberger 2002, van Rooij 2004) and it contains Bar-Hillel's and Carnap's theory of semantic information (Bar-Hillel and Carnap 1953, 1964) as a special case. The application of the theory to examples from computational mechanics (Crutchfield 1991, 1992), to the determination of pragmatic information due to efficiency changes (Gernert 1985, Atmanspacher and Scheingraber 1990, Crutchfield 1992) is left open for future research.

Further developments might combine Gärdenfors' revision-extended Bayesian belief models (Gärdenfors 1988), which actually provide instances of propositional (or intuitive) logics with more elaborated accounts (Staudacher 1987, Groenendijk and Stokhof 1991, Kracht 2002) supplying the full power of predicate logics. In this prospect, the proposed theory of pragmatic information could become a promising tool for relevance theory and experimental pragmatics.

Acknowledgments

The research work for this paper was mainly conducted during a visit at the Institute for Frontier Areas of Psychology, Freiburg, in July 2004.

I thank Harald Atmanspacher for his kind hospitality and for valuable discussions. I am also indebted to Peter Staudacher, Reinhold Kliegl, Douglas Saddy, Dieter Gernert, Thomas Weskott, Robin Hörnig, and two referees for their critical comments who helped me to substantially improve the manuscript. I gratefully acknowledge financial support from the *Deutsche Forschungsgemeinschaft* to the former center of excellence "Formal Models of Cognitive Complexity" and to the research group "Conflicting Rules in Cognitive Systems", both at the University of Potsdam.

References

Atmanspacher H. and Filk T. (2006): Complexity and non-commutativity of learning operations on graphs. *BioSystems* **85**, 84–93.

Atmanspacher H., Kurths J., Scheingraber H., Wackerbauer W., and Witt A. (1992): Complexity and meaning in nonlinear dynamical systems. *Open Systems & Information Dynamics* **1**, 269–289.

Atmanspacher H., Räth C., and Wiedenmann G. (1997): Statistics and metastatistics in the concept of complexity. *Physica A* **234**, 819–829.

Atmanspacher H., Römer H., and Walach H. (2002): Weak quantum theory: Complementarity and entanglement in physics and beyond. *Foundations of Physics* **32**, 79–406.

Atmanspacher H. and Scheingraber H. (1990): Pragmatic information and dynamical instabilities in a multimode continuous-wave dye laser. *Canadian Journal of Physics* **68**, 728–737.

Atmanspacher H. and Wiedenmann G. (1999): Some basic problems with complex systems. In *Large Scale Systems: Theory and Applications*, ed. by N. T. Koussoulas and T. Groumpos, Elsevier, Amsterdam, pp. 1059–1066.

Badii R., Finardi M., and Broggi G. (1991): Recursive symbolic encoding and hierarchical description of complex systems. In *Information Dynamics*, ed. by H. Atmanspacher and H. Scheingraber, Plenum, New York, pp. 35–43.

Badii R. and Politi A. (1997): *Complexity. Hierarchical Structures and Scaling in Physics*, Cambridge University Press, Cambridge.

Bar-Hillel Y. and Carnap R. (1953): Semantic information. *British Journal for the Philosophy of Science* **4**, 147–157.

Bar-Hillel Y. and Carnap R. (1964): An outline of a theory of semantic information. In *Language and Information*, ed. by Y. Bar-Hillel, Addison-Wesley, Reading MA, pp. 221–274.

Bernardo J.M. (1979): Expected information as expected utility. *Annals of Statististics* **7**, 686–690.

Chaitin G.J. (1977): Algorithmic information theory. *IBM Journal for Research Development* **21**, 350–359.

Crutchfield J.P. (1991): Reconstructing language hierarchies. In *Information Dynamics*, ed. by H. Atmanspacher and H. Scheingraber, Plenum, New York, pp. 45–59.

Crutchfield J.P. (1992): Semantics and thermodynamics. In *Nonlinear Modeling and Forecasting*, Addison-Wesley, Reading MA, pp. 317–359.

Crutchfield J.P., Feldman D.P., and Shalizi C.R. (2000): Comment I on "simple measure for complexity". *Physical Review E* **62**, 2996–2997.

Crutchfield J.P. and Young K. (1989): Inferring statistical complexity. *Physical Review Letters* **63**, 105–108.

Crutchfield J.P. and Young K. (1990): Computation at the onset of chaos. In *Complexity, Entropy and the Physics of Information*, ed. by W.H. Zurek, Addison-Wesley, Reading MA, pp. 223–269.

Frank A.U. (2003): Pragmatic information content – how to measure the information in a route description. In *Perspectives on Geographic Information Science*, ed. by M. Goodchild, M. Duckham, M. Worboys, Taylor and Francis, London, pp. 47–68.

Frisch S., beim Graben P., and Schlesewsky M. (2004): Parallelizing grammatical functions: (P600) and (P345) reflect different cost of reanalysis. *International Journal of Bifurcation and Chaos* **14**, 531–549.

Fu K.S. (1974): *Syntactic Methods in Pattern Recognition*, Academic Press, New York.

Gärdenfors P. (1988): *Knowledge in Flux. Modeling the Dynamics of Epistemic States*, MIT Press, Cambridge MA.

Gärdenfors P. (1994): How logic emerges from the dynamics of information. In *Foundations of Computation*, ed. by J. v. Eijck and A. Visser, MIT Press, Cambrigde MA, pp. 49–77.

Gernert D. (1985): Measurement of pragmatic information. *Cognitive Systems* **1**(3), 169–176.

Gernert D. (1996): Pragmatic information as a unifying concept. In *Information. New Questions to a Multidisciplinary Concept*, ed. by K. Kornwachs and K. Jacoby, Akademie, Berlin, pp. 147–162.

Gernert D. (2000): Towards a closed description of observation processes. *BioSystems* **54**, 165–180.

Gernert D. (2006): Pragmatic information: Historical development and general overview. *Mind and Matter* **4**, 141–167.

Goodman N. (1983): *Fact, Fiction and Forecast*, Harvard University Press, Cambridge MA.

Grassberger P. (1986): Toward a quantitative theory of self-generated complexity. *International Journal of Theoretical Physics* **25**, 907–938.

Grassberger P. (1989): Problems in quantifying self-generated complexity. *Helvetica Physica Acta* **62**, 489–511.

Grassberger P. (1991): Information and complexity measures in dynamical systems. In *Information Dynamics*, ed. by H. Atmanspacher and H. Scheingraber, Plenum, New York, pp. 15–33.

Groenendijk J. and Stokhof M. (1991): Dynamic predicate logic. *Linguistics and Philosophy* **14**, 39–100.

Kolmogorov A.N. (1968): Logical basis for information theory and probability theory. *IEEE Transactions on Information Theory* **14**, 662–664.

Kornwachs K. and Lucadou W. von (1982): Pragmatic information and non-classical systems. In *Cybernetics and Systems Research*, ed. by R. Trappl, North Holland, Amsterdam, pp. 191–197.

Kornwachs K. and Lucadou W. von (1985): Pragmatic information as nonclassical concept to describe cognitive processes. *Cognitive Systems* **1**(1), 79–94.

Kracht M. (2002): Dynamic semantics. *Linguistische Berichte*, Sonderheft X, 217–241.

Kullback S. (1968): *Information Theory and Statistics*, Dover, New York.

Kullback S. and Leibler R.A. (1951): On information and sufficiency. *Annals of Mathematical Statistics* **22**, 79–86.

Morris C.W. (1955): Foundations of the theory of signs. In *International Encyclopedia of Unified Science Vol. I.2*, ed. by O. Neurath, R. Carnap, and C.W. Morris, University of Chicago Press, Chicago, pp. 77–137.

Polani D., Martinetz T., and Kim J. (2001): An information-theoretic approach for the quantification of relevance. In *Advances in Artificial Life*, ed. by J. Kelemen and P. Sosih, Springer, Berlin, pp. 704–713.

Quian Quiroga R., Arnhold J., Lehnertz K., and Grassberger P. (2000): Kullback-Leibler and renormalized entropies: Applications to electroencephalograms of epilepsy patients. *Physical Review E* **62**, 8380–8386. See also their reply to a comment on their paper in *Physical Review E* **66**, 043903 (2002).

Saparin P., Witt A., Kurths J., and Anishchenko V. (1994): The renormalized entropy – an appropriate complexity measure? *Chaos, Solitons & Fractals* **4**, 1907–1916.

Shalizi C.R., Shalizi K.L., and Haslinger R. (2004): Quantifying self-organization with optimal predictors. *Physical Review Letters* **93**, 118701.

Shannon C.E. and Weaver W. (1949): *The Mathematical Theory of Communication*, University of Illinois Press, Urbana.

Shiner J.S., Davison M., and Landsberg P.T. (1999): Simple measure for complexity. *Physical Review E* **59**, 1459–1464.

Staudacher P. (1987): Zur Semantik indefiniter Nominalphrasen. In *Neuere Forschungen zur Wortbildung und Historiographie der Linguistik*, ed. by B. Ansbach-Schnitker and J. Roggenhofer, Gunter Narr, Tübingen, pp. 239–258.

Römer H. (2004): Weak quantum theory and the emergence of time. *Mind and Matter* **2**(2), 105–125.

Rooij R. van (2004): Utility, informativity and protocols. *Journal of Philosophical Logic* **33**, 389–419.

Rooij R. van (2006): Pragmatic value and complex sentences. *Mind and Matter* **4**, 195–218.

Weizsäcker C.F. von (1974): *Die Einheit der Natur*, Hanser, München.

Weizsäcker C.F. von (1985): *Aufbau der Physik*, Hanser, München.

Weizsäcker E.U. von (1974): Erstmaligkeit und Bestätigung als Komponenten der pragmatischen Information. In *Offene Systeme I*, ed. by E. U. von Weizsäcker, Klett-Cotta, Stuttgart, pp. 82–111.

Weizsäcker E.U. von and Weizsäcker C. von (1972): Wiederaufnahme der begrifflichen Frage: Was ist Information? *Nova Acta Leopoldina* 37/1 (206), 536–555.

Weizsäcker E.U. von and Weizsäcker C. von (1998): Information, evolution and "error-friendliness". *Mind and Matter* **4**, 235–247.

Wackerbauer R., Witt A., Atmanspacher H., Kurths J., and Scheingraber H. (1994): A comparative classification of complexity measures. *Chaos, Solitons & Fractals* **4**, 133–173.

Weinberger E.D. (2002): A theory of pragmatic information and its application to the quasi-species model of biological evolution. *BioSystems* **66**, 105–119.

Williams P.M. (1980): Bayesian conditionalization and the principle of minimum information. *British Journal for the Philosophy of Science* **31**, 131–144.

Received: 13 February 2006
Revised: 10 July 2006
Accepted: 13 October 2006

Reviewed by two anonymous referees.