Conference Objectives

Info-metrics is the science and practice of inference and quantitative information processing. It is applicable to all sciences and provides the universal mathematical and philosophical foundations for inference with finite, noisy or incomplete information.

The objectives of this conference are to continue the study of some of the open questions within philosophy of information with an emphasis on the study of the value of information. Interest in the philosophy and meaning of information goes back half a century but has rapidly increased recently with many new directions of research into the meaning, quantification and measures of information and complexity as well as a vast range of applications across the scientific spectrum.

In our first workshop on the topic (October, 2011) we focused on one aspect of the philosophy of information: the different techniques to measure information and to identify meaningful information.

In my opening remarks, I provided a background on info-metrics, and then provided a fast review of the philosophy of information with an emphasis on the value of information (both relative and “absolute”). I also provided a new theorem on the “absolute” value of information. Briefly stating, in this workshop we focused on a number of related aspects of the philosophy of information and the philosophy of info-metrics. These topics include the value of information, quantifying information, processing complementary and contradicting information and the inter-relationship between information, computation and complexity.

Specific papers included information and its value in big data, information as a good, contradictory information and its value, beliefs and desire and their affect on the subjective value of information, visualization of information and ignorance and information.
James Moor (Dartmouth)

The Dark Side of Big Data

Although there are many advantages in using big data techniques, there are dangers as well. This paper examines some of the challenges to our privacy, autonomy, and responsibility.
Luciano Floridi (Hertfordshire and Oxford)

Maker’s Knowledge and the Synthetic Uninformative

In this presentation, I begin by discussing the three standard distinctions used to qualify propositional knowledge: analytic vs. synthetic, a priori vs. a posteriori, and necessary vs. contingent. The ultimate goal is to understand what kind of knowledge is the so-called maker’s knowledge, as when Alice (knows or rather) is informed (holds the information) that Bob’s coffee is sweetened because she just put two spoons of sugar in it. In the course of the presentation, I shall argue that:

a) we need to decouple a fourth distinction, namely informative vs. uninformative, from the previous three, in particular from its implicit association with analytic vs. synthetic and/or a priori vs. a posteriori;

b) such decoupling facilitates, and is facilitated by, moving from a propositional to an agent-oriented approach: the distinctions qualify a proposition, a message, or a set of well-formed, meaningful and truthful data not just in themselves but with respect to an information agent;

c) the decoupling and the agent-oriented approach enable a re-mapping of currently available positions (Classic, Innatist, Kant’s and Kripke’s) on these four dichotomies; and

d) within such a re-mapping, a fifth position, capturing the nature of a maker’s information in terms of these four dichotomies, is best described as the synthetic uninformative.
**J. Michael Dunn (Indiana U. Bloomington)**

**Contradictory Information Can Be Valuable: The Paradox of the Two Firemen**

This is a follow-up to my paper "Contradictory Information: Too Much of a Good Thing" (Dunn (2009)), in which I embedded the “Belnap-Dunn 4-valued Logic” (Truth, Falsity, Neither, Both) into a context of subjective probability generalized to allow for degrees of belief, disbelief, and uncertainty, with the third value split into two kinds of uncertainty— that in which the reasoner has too little information (ignorance) and that in which the reasoner has too much information (conflict). Jøsang’s (1997) “Opinion Triangle” was thus expanded to an “Opinion Tetrahedron” with the 4-values as its vertices.

Floridi (2011) develops what he calls a "strong semantic theory of information" that avoids what he labels the Bar-Hillel Carnap Paradox (that contradictions contain an infinite amount of information)." Floridi says contradictions contain zero information, giving among other reasons that "inconsistent information is obviously of no use to a decision maker."

Suppose you are awakened in your hotel room by a re alarm. You open the door. You see three possible ways out: left, right, straight ahead. Scenario 1. You see two firemen. One says there is exactly one safe route and it is to your left. The other says it is to your right. Contradictory information! Scenario 2. You find no one to give directions. Incomplete information!

I think it is obvious, other things being equal, that a rational agent would prefer to be in Scenario 1. While the two firemen are giving you contradictory information, they are also both giving you the useful information that there is a safe way out and it is not straight ahead. Three choices have been reduced to two, thereby increasing your odds for survival. I assume you have every reason to believe that the firemen are experts, good willed, with their recommendations evidence-based. If only one of them had made his statement you would have followed his advice without hesitation.

I consider possible reactions to The Two Firemen, and defend it as being a clear example of how contradictory information can sometimes be useful information that is relevant to decision making (contrary to Floridi). I will also discuss other examples in the context of unstructured databases (think WWW). I believe the Opinion Tetrahedron can adequately evaluate the firemen’s claims, but others may have their own "ways out."

**References**


**Pieter Adriaans (U. Amsterdam)**

**Does information have intrinsic value?**

In philosophy we have the distinction between intrinsic and instrumental values. Human beings or works of art may be considered to have ‘intrinsic’ value in themselves. Intrinsic values also have been associated with Plato’s idealism: the world of ideas is one of eternal beauty and truth and has value independent of any human. A dollar bill on the other hand only has value because of rules and conventions we have agreed on. The value of money is instrumental in so far as it helps us to reach certain goals. There is no doubt that information can have instrumental value. If I go to the Tourist Office in a strange city and ask for directions to a hotel, the map I get really contains information that is ‘instrumental’ in helping me to get there.

So can we make a case for intrinsic value of information? Of course it depends on how we measure information. According to Shannon’s notion of information the amount of information is dependent on its probability. The value of a message then could only be intrinsic if we have a way of assigning an intrinsic probability to it. The theory of Kolmogorov complexity and its notion of a Universal Distribution tries to do this. It helps us to understand the notion of randomness, and, in fact, a random string can be said to have some intrinsic value. Firstly, it is impossible to prove that a string is really random, secondly we can use random strings to construct absolutely undecipherable codes.

Yet, a drawback of all these measures is that they assign the highest information content to data sets with the highest entropy and that is not what we intuitively would see as valuable: e.g. it would mean that a television broadcast with white noise has the most value. In the past decennium another theory that goes under the name of two-part code optimization has caught a lot of attention. One of its formulations is the so-called Minimum Description Length (MDL) principle: The best theory to explain a set of data is the one which minimizes the sum of: the length, in bits, of the description of the theory and the length, in bits, of the data when encoded with the help of the theory. The MDL principle balances the amount of model information with the amount of ad hoc data in a data set.

With the history of philosophy in mind, the amount of model information in a data set seems to be a good candidate for an estimate of the intrinsic value. An interesting historical parallel is the association of the notion of model with that of a platonic form or idea. In a universe in which everything in the end reaches a state of maximum entropy a structure that manages to maintain its integrity or form over time, whether it is a human being, a work of art, or a written meaningful message seems to be something that is intrinsically valid, independent of the fact whether there exists any agent to appreciate this value. Given the fact that the term in-form-ation was coined by authors like Cicero and Augustine as a translation of the term like ‘idea’ and ‘form’ from Greek there is more historical continuity in the study of information than one might have thought.
Belief and Desire: On Information and its Value

The challenge of reasoning under conditions of uncertainty is the single common thread that unifies all sciences. In this talk I will argue that the challenge can be successfully met by addressing three interrelated problems.

The first problem is to design a scheme that allows one to represent a state of partial knowledge. The result is well known—it is given by Bayesian probability theory. Even though the foundations of Bayesianism remain an active subject of discussion, the practical success of the whole framework is by now undeniable.

The second problem is to design a procedure that allows us to revise our beliefs as we acquire new information. Indeed, if the source of all our troubles is incomplete information what do we do in the fortunate circumstance that some information becomes available? Which raises another question: what, after all, is information? I shall argue that in a Bayesian setting the notion of information must be defined in terms of its effects on the beliefs of rational agents: Information is what induces a rational agent to change its mind. The tool to perform the updating is entropy and the associated method, Entropic Inference, includes the MaxEnt method and Bayes' rule as special cases.

The actual functional form of the entropy follows from purely pragmatic requirements. The method must be of universal applicability and we must recognize that prior information is a valuable asset that should not be squandered. Which brings us to the third problem: Why bother? What do we expect to gain by going through the trouble of collecting and processing information? The answer is that we are driven not just by what we believe but also by what we desire. Better information leads to better beliefs which lead to decisions that stand a better chance of getting us what we want. Thus the value of information is to be measured in terms of differences in the expected utility of the decisions we make.

Caticha’s papers on entropic inference and on its applications to the foundations of statistical and quantum mechanics can be found at http://www.albany.edu/physics/acaticha.shtml.
Duncan Foley (New School for Social Research and Santa Fe Inst.)

These notes are a continuation of a long and productive conversation with David Oliver on the sources of randomness in the natural world. I’d like to acknowledge helpful conversations with Murray Gell-Mann and Seth Lloyd, and helpful comments on earlier versions from Gregor Semeniuk and Ellis Scharfenaker.

This paper examines the foundations of the concept of effective complexity proposed by Murray Gell-Mann and Seth Lloyd using the methods of Bayesian inference. Given a data string, \( x \in \{0, 1\}^\infty \) of length \( |x| \), an ensemble, \( \mathcal{E} = \{X_\mathcal{E}, \omega_\mathcal{E}\} \), assigns a frequency, \( \mathcal{E}[x'] \) to strings \( x' \) with \( |x'| = |x| \) of the same length as the data string. Programs to compute ensemble frequencies partition (or coarse-grain) the set of strings into exclusive subsets \( X_\mathcal{E} = \{X_\mathcal{E}^1, \ldots, X_\mathcal{E}^r\} \), and assign a coarse-grained frequency distribution, \( \omega_\mathcal{E} = \{\omega_\mathcal{E}^1, \ldots, \omega_\mathcal{E}^r\} \), \( \omega_\mathcal{E}^r \geq 0 \), \( \sum_1^r \omega_\mathcal{E}^r = 1 \) to the subsets. Let \( f(x') \) be the index of the subset containing the string \( x' \). The ensemble frequency of every string \( x' \in X_\mathcal{E}^r \) is \( \mathcal{E}[x'] = \omega_\mathcal{E}^r / |X_\mathcal{E}^r| \), where \( |X_\mathcal{E}^r| \) is the number of strings in the subset \( X_\mathcal{E}^r \). According to Bayes’ theorem, given a prior assignment of probabilities over a class of ensembles, such as the class of computable ensembles, \( P[\mathcal{E}] \), the log posterior probability of any ensemble given the data string is \( \log[P(\mathcal{E})] + \log[P(x | \mathcal{E})] \), the sum of the log of the ensemble’s prior probability and the log of the data string’s conditional probability given the ensemble (its likelihood in Bayesian terminology). The ensemble specifies the location and coarse-grained frequency of the subset containing the data string, \( \omega_{f(x)} \), so we have \( \log[P(x | \mathcal{E})] = -\log[|X_\mathcal{E}^f|] \). An ensemble’s Kolmogorov algorithmic information content (AIC), \( K[\mathcal{E}] \), is the length of the shortest program string that leads a prefix-free universal Turing machine to compute the ensemble frequencies when the string and a precision \( n \) are presented to it as input. With a universal prior over ensembles that is inversely proportional to Kolmogorov algorithmic information content, \( \log[P^U[\mathcal{E}]] = -K[\mathcal{E}] \), the log posterior probability of an ensemble given the data string is \( -K[\mathcal{E}] + \log[P(x | \mathcal{E})] \), the sum of the negative of the ensemble AIC and the frequency of the data string under the ensemble, this sum is also the negative of the length of the two-part code for the data string using the ensemble as the model. Energy-like constraints may limit the set of admissible ensembles. The AIC of an ensemble is the sum of the AIC of the partition and the bits required to specify the coarse-grained frequencies, \( K[\mathcal{E}] = K[X_\mathcal{E}] + K[\omega_\mathcal{E}] \), the total information of the ensemble. The term \( K[\omega_\mathcal{E}] \) depends on the prior distribution over frequency distributions that corresponds to the coding system used to specify the frequency distribution. If, for example, the prior probability of a coarse-grained frequency distribution is inversely proportional to its entropy relative to the uniform distribution, the Shannon code length for a coarse-grained frequency distribution is \( K[\omega_\mathcal{E}] = \log[|\omega_\mathcal{E}|] - H[\omega_\mathcal{E}] \). With this coding of ensemble frequencies the log posterior probability of an ensemble is \( -K[X_\mathcal{E}] - \log[|\omega_\mathcal{E}|] + H[\omega_\mathcal{E}] - \log[|X_\mathcal{E}^f|] \). For any partition of the strings, with this particular coding of frequency distributions maximization of the posterior probability leads to a maximum entropy coarse-grained frequency distribution subject to the constraints. Other coding systems can result in minimum entropy frequency distributions or frequency distributions that maximize or minimize other entropy-like quantities. In general partitions that put the data string in a small subset require more subsets, a longer program, and a lower coarse-grained entropy to meet the constraints.
Min Chen (Oxford)

Schools of Thought in Visualization

Quality of visualization (QoV) has always been the main catalyst that motivates research and development in visualization. However, whenever we start to discuss QoV, the term becomes elusive. We often appear to have diffident emphases. In this talk, the speaker will examine the four schools of thoughts that were presented in an IEEE VisWeek 2012 panel, entitled "Quality of Visualization: The Bake Off". These four schools of thought are A for Algorithms and Automation; E for Empirical Studies and Experiments; M for Metrics and Measurements; and R for Real Users and Real World Applications. In the era of data deluge, the discussion on QoV is timely. It also signifies the scholarly transformation of the field of visualization towards a mature discipline.
Teddy Seidenfeld (Carnegie Mellon)

Is Ignorance Bliss for Sets of Probabilities?

When uncertainty is represented with sets of probabilities -- so called "Imprecise Probability" [IP] theory -- rather than with a single probability distribution, what becomes of the basic result from Subjective Expected Utility [SEU] theory that cost-free information has non-negative value? That is, under SEU theory, where uncertainty is represented by a single probability distribution and the decision rule is to maximize expected utility -- and in the absence of moral hazard-- the rational agent, YOU, should postpone a decision in order first to acquire new, cost-free information. SEU theory advises YOU to value cost-free information as having non-negative instrumental value for making better decisions. If the new information might alter your decision, then ex-ante, such information has positive value to YOU. Only if YOU will decide the same, regardless of the new information, it has no ex-ante added value for YOU.

I review necessary and sufficient conditions, within IP theory, for new evidence to dilate a set of probabilities. Evidence from an experiment dilates a set of probabilities when, for each possible experimental outcome, the uncertainty in the updated IP set of probabilities increases. The IP increase in uncertainty happens because the lower and upper conditional probabilities, given the new evidence, are more extreme than their corresponding lower and upper unconditional probabilities. That is, with dilation the new evidence is certain to increase IP-uncertainty in the sense that the updated IP intervals of probability become strictly larger compared with the IP intervals of probability prior to updating.

For a variety of IP-decision rules, an experiment that causes dilation produces cost-free information with negative IP value. For instance, with dilation, the IP decision rule to maximize minimum expected utility leads YOU to give the new evidence negative value: YOU strictly prefer to make the decision without updating first, contrary to the basic SEU result summarized above.

The same results about dilation can be interpreted as showing when new, shared information among several Bayesians is sure to lead them to diverging posterior probabilities, in contrast with the familiar (asymptotic) results about the merging of posterior probabilities with shared evidence. With dilation, the priors swamp the data in the posteriors, and not the other way around!