Spontaneous assessment of complexity in the selection of events

Calcul spontané de la complexité dans la sélection des événements

Jean-Louis Dessalles

2007D011

Août 2007

Département Informatique et Réseaux
Groupe IC2 : Interaction, Cognition et Complexité

Spontaneous assessment of complexity in the selection of events

Calcul spontané de la complexité dans la selection des événements

Jean-Louis Dessalles

ParisTech – ENST (LTCI, CNRS UMR 5141)
46 rue Barrault – 75013 Paris, France
dessalles@enst.fr - www.enst.fr/~jld

Most of the situations of daily life that arouse human interest are experienced as *unexpected*. Highly unexpected events are preferentially memorised and are systematically signalled or reported in conversation. Probability theory is shown to be inadequate to predict which situations will be perceived as unexpected. We found that unexpectedness is best explained using Kolmogorov complexity, which is a strong indication that human individuals have an intuitive access to what was thought to be only an abstract mathematical notion. Many important and previously disparate facts about human communicative behaviour are shown to result from the cognitive ability to detect complexity shifts.

La plupart des situations de la vie quotidienne qui retiennent l'intérêt des êtres humains sont perçues comme inattendues. Les événements particulièrement inattendus sont préférentiellement mémorisés et sont systématiquement rapportés lors de conversations. La théorie des probabilités se révèle inadéquate pour prédire quelles situations seront perçues comme inattendues. Notre étude montre que l'inattendu est expliqué par un saut de la complexité de Kolmogorov, ce qui suggère que les individus humains ont un accès intuitif à ce qui pouvait apparaître comme une notion mathématique abstraite. Des propriétés importantes et précédemment disparates relatives au comportement communicationnel humain se trouvent résulter de la capacité cognitive à détecter les sauts de complexité.

Spontaneous assessment of complexity in the selection of events

Jean-Louis Dessalles

ParisTech – ENST (LTCI, CNRS UMR 5141)
46 rue Barrault - 75013 Paris, France
dessalles@enst.fr - *www.enst.fr/~jld*

Most of the situations of daily life that arouse human interest are experienced as *unexpected*. Highly unexpected events are preferentially memorised and are systematically signalled or reported in conversation. Probability theory is shown to be inadequate to predict which situations will be perceived as unexpected. We found that unexpectedness is best explained using Kolmogorov complexity, which is a strong indication that human individuals have an intuitive access to what was thought to be only an abstract mathematical notion. Many important and previously disparate facts about human communicative behaviour are shown to result from the cognitive ability to detect complexity shifts.

Human beings perform a daily task that, as far as we know, no other animals do. They devote a considerable amount of their time reporting events, as the observation of their conversational behaviour shows[1]. They select among their various experiences those which are worth reporting to conspecifics. Only a small fraction passes the selection. Though reported experiences may belong to any domain of experience, there are strong requirements for a state of affairs to become an interesting event. One of them is that it be *unexpected*. The characterisation of unexpectedness reveals that human individuals actively estimate the complexity of situations, in the Kolmogorov sense.

Several parameters are known to systematically influence the unexpectedness and thus the interest of an occurring event, among which rarity, atypicality and proximity. The reportability of a robbery grows if robberies are rare in the region, if the stolen property was atypical by its high amount and if it concerned one's neighbour. Proximity in space[2,3], in time[4] and departure from norms[5,6] have been shown to have a systematic influence on interest and newsworthiness.

One natural way to unify these apparently unrelated parameters is to say that the unexpectedness of an event is measured by its *improbability*, thus implementing an information-theoretic approach to human interest. Shannon's definition of transmitted information[7] can be applied to event-related communication between humans by defining unexpectedness as $U = \log_2 1/p_i$, where p_i is the *a priori* probability of the event. An alternative definition[8] is $U = \Sigma\, p_j^2/p_i$. It accounts for the fact that improbable events are only interesting if they are a contrast to probable alternatives. These definitions based on probability lead however to considerable difficulties.

Probability does not account for unexpectedness

The first problem is that human beings are known to perform badly when combining probabilities. Various biases, such as the gambler's fallacy[9], the conjunction fallacy[10], the base-rate 'fallacy'[11,12], the over-sensitivity to personal data[13], to representativeness[14] and to cognitive availability[15] lead individuals to make glaring but consistent judgment errors even in simple situations. By inverting the perspective, probability theory can be claimed to offer a poor image of human probabilistic judgment.

A second problem comes from the artificial conflation, under the same concept of probability, of apparently unrelated sources of unexpectedness: rarity and unexpected proximity (a robbery in one's neighbour's house), unexpected deviation (a dog of the size of a rat), or unexpected coincidences (getting five ones when rolling five dice). This conflation is achieved through abstract mathematical concepts such as Poisson or Gaussian distribution laws and measure theory, which allow to establish a link between continuous and discrete probability. It is unlikely that human minds have any intuitive access to these concepts and use them to get a unified notion of unexpectedness.

A third problem with a probability-theoretic approach to unexpectedness is that most definitions of probability, including the axiomatic, the frequentist, the subjective and Bayesian frameworks, require that the set of exclusive alternatives be given in advance, whereas it is rarely available in ordinary life. *Each actual state of affairs is unique*. Probability theories provide no general rule whatsoever indicating which aspects, such as the colour or the precise position of the dice on the table, should be ignored when defining what is to be considered the actual event and what its alternatives are.

A fourth problem comes from the fact that individuals regard as unexpected situations that *qualitatively* differ from prototypes, such as a towing attachment on an expensive sport car. Probability theory is unable to predict the value of unexpectedness in such case: subjects have no access to any statistics that might include sport cars with such equipments.

In what follows, we give a formal definition of unexpectedness based on complexity. We then show that the model makes correct predictions about crucial factors that control interest (Table 1). We finally discuss the plausibility of complexity assessment by humans, before mentioning potential applications of the model.

Unexpectedness as Kolmogorov complexity drop

The quest for a better model for determining factors of interest was naturally oriented toward the notion of complexity when considering examples as the following. Encountering by chance a close acquaintance in a remote place can be highly unexpected. Unexpectedness in this case is an increasing function of the remoteness of the place and of the friend's closeness. In other words, for the story to be reportable, the place has to be *complex* whereas the friend has to be *simple*. Probability theory is of little help to explain the gradual influence of complexity on unexpectedness in such an example. Another example can be taken from lottery drawings[16]. Subjects consider the occurrence of simple drawings such as 1-2-3-4-5-6 as virtually or even strictly impossible, whereas complex configurations such as 12-27-31-36-37-41 seem worth wagering. Correlatively, the public does not care if the latter is actually drawn, whereas the occurrence of the former would make formidable news. Again, probability theory does not account for the differential unexpectedness of these drawings.

Table 1 – Some predictions of the model

Prediction	Example	Relevant parameters
Deviant feature	*Sport car with a towing attachement*	Distinctiveness of the deviant object Simplicity of the deviant feature
Remarkable structure	*Consecutive lottery drawing*	Simplicity of the structure
Coincidence	*Two unrelated but similar suicides by drowning*	Strength of the analogy
Proximity	*Fire in the neighbourhood*	Short distance
Landmarks	*Blaze on the Eiffel Tower*	Simplicity of the landmark
Topic relatedness	*Car accessories introducing towing attachment*	Strength of the analogy
Encounter problem	*Encountering a know person in a remote place*	Complexity of the place Simplicity of the person

We discovered that Kolmogorov complexity provides a much better measure of what human beings consider unexpected. The Kolmogorov complexity $K(s)$ of a state of affairs s is the size of the shortest programme that generates s if run on a universal computing machine[17]. This abstract definition has two major drawbacks when applied to finite objects: it depends on the chosen machine, and it is generally not computable. The concept of complexity is less problematic when applied to cognitive processes[18]. We introduce the notion of *cognitive complexity* $C(\ )$ as an instantiation of $K(\ )$ defined for a specific 'machine', the observer's mind. $C(s)$ is defined as *the length of a minimal cognitive procedure through which the observer can generate the state of affairs* s. Thanks to this definition, we could characterize unexpectedness and reach a predictive model of interest (Table 1).

Though cognitive complexity may remain out of reach in many cases, complexity differences are most often computable. This makes unexpectedness an operational concept. A situation is *unexpected* if taking into account some feature F generates a complexity drop.

$$U(s,F) = C(s) - C(F) - C(s|F) \tag{1}$$

F may be any detail mentioned when an event is recounted. $C(s|F)$ is the complexity of generating *s* when *F* is already available.[1] A feature *F* is said *relevant* to *s* if $U(s,F) \geq 0$, otherwise it is irrelevant. Only relevant features are considered here. $C(s)$ and $C(s|F)$ can be dubbed *a priori* and *a posteriori* complexity. They are relative to the same event, but correspond to different computations. The computation of $C(s)$ relies on a standard procedure available to the subject. It is generally based on a prototypical situation *r* to which *s* can be associated. By comparison, the generation of $C(s|F)$ can be reduced, as *F* may provide a shortcut in the computation of *s* (figure 1). Unexpected situations are simpler *a posteriori* than *a priori*. Definition (1) illustrates the fact that simplicity plays a fundamental role in cognition[19], even at the highest level where individuals determine their focus of interest.

For example, if *s* designates a given expensive sport car, the standard complexity of *s* amounts to the minimum number of yes/no questions necessary to specify the actual car among all imaginable sport cars. It is expected to be $\log_2 N$ if *N* is the number of sport cars. If that car happens to carry a towing attachment (*F*) that is believed to make it unique in its kind (as towing and speeding seem contradictory), then $C(s|F)$ is zero and $U(s,F) = \log_2 N - C(F)$. An important prediction here is that distinguishing features are most unexpected when they are *simple* (figure 2).

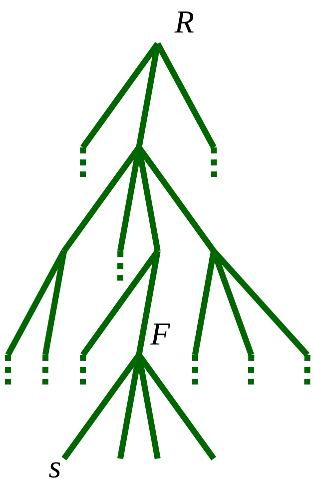

Figure 1: **A simplified tree model**
In this simple model, a situation *s* is conceived as resulting from a series of choices leading from a starting point *R* to *s* through various intermediary states (*e.g.* choosing a path when visiting a foreign city). The complexity of *s* is the number of bits necessary to specify the path from the root *R* to *s*. It amounts to $\log_2 3 + \log_2 4 + \log_2 2 + \log_2 4$ in the example. $C(s|F)$ refers to the complexity of the path from *F* to *s* and amounts to $\log_2 4$ in the example. The difference $C(s) - C(s|F)$ can be assessed even if one has only partial information about *s*. In this tree model, it amounts to the portion of the path leading to *s* that stops at *F*. If the only access to *F* is that partial path, then $U(s,F)$ is zero. If *F* can be acceded to though some other way, *e.g.* if nodes are independently listed and *F* turns out to be the first node in that list, then $C(F)$ is minimal and $U(s,F)$ reaches a significant value. Note that the unexpectedness of *s* can be assessed without knowing its alternative, which would be the leaves in this simple tree model.

For similar reasons, each banknote is unique thanks to the number it carries, but only simple numbers such as 121212121212 or 123456654321 can make them *unexpected* and thus noteworthy.[2] These numbers can be generated by combining simple operations such as increment, copy, symmetry, so that they can be represented

---

1 This definition of conditional complexity matches the standard one, which supposes that *F* is given as input to the procedure that produces *s*.

2 In some definitions of complexity-based probability[20], simple structures are mostly probable. Here, on the contrary, the most unexpected structures are those that offer a complexity contrast by being simpler than expected.

by a hierarchy of nested group structures[21]. Their representation in a human mind is much more concise than numbers like 491944264709 that seem devoid of structure. In this context, the unexpectedness of a number constrained by a repetitive structure, like 333333333333, is $11*\log_2 10 = 36.54$ bits, as it requires one instantiation and a copy instead of a copy followed by 12 instantiations.

Figure 2: **Simplicity of distinguishing features**
This 24-cent airmail stamp of 1918, which was erroneously printed with an inverted centre, is worth about $200,000, about two thousand times the price of a regular copy of the same stamp. The interest in such an item is enhanced by the structural simplicity of what makes it particular (here a mere inversion), as suggested by our definition of unexpectedness.

(from the Smithsonian National Postal Museum, www.postalmuseum.si.edu/exhibits/2f1a_inverts.html)

Coincidences and unexpected simplicity

Coincidences exert special attraction on human minds[22]. For instance, a news item like two persons having committed suicide by drowning was reported in the French national media[23] because of the striking similarity of the two events. Cognitive complexity provides a straightforward explanation why the co-occurrence of two similar events elicits high interest. Assuming that e_1 and e_2 are two independent (*i.e.* non causally related) events, the standard computation of $C(e_1 \& e_2)$ amounts to $C(e_1) + C(e_2)$, and:

$$U(e_1 \& e_2, e_1) = C(e_2) - C(e_2|e_1) \tag{2}$$

Equation (2) predicts high unexpectedness when the co-occurring events bear an analogy to each other, as analogy minimizes complexity[24]. Unexpectedness also rises with the complexity of the coinciding events. Hence the importance, when reporting coincidences, of mentioning every detail that makes the two events more particular while preserving their similarity, as in the well-known parallel between the lives of A. Lincoln and J. F. Kennedy[25].

The notion of cognitive complexity also explains egocentric effects in the perception of coincidences, which bring subjects to be more surprised at coincidences happening to them than to comparable coincidences happening to others, even when the latter is objectively more surprising[26]. For instance, a match of birthdays in a group appears more surprising to subjects who are involved in it[21]. To represent this effect, we introduce *computation sequences* (noted with operator +) that provide indications for the actual computation of complexity:

$$U(s, F_1+F_2) = C(s) - C(F_1) - C(F_2|F_1) - C(s|F_1 \& F_2) \tag{3}$$

In the frequent case in which the computation sequence is memoryless, *i.e.* $C(s|F_1 \& F_2) = C(s|F_2)$, we can derive:

$$U(s, F_1+F_2) = U(s, F_2) + U(F_2, F_1) \quad (4)$$

In the case of coincidences, the sequence $ego+e_1$ is memoryless, since $C(e_1\&e_2|ego\&e_1) = C(e_2|ego\&e_1) = C(e_2|e_1)$, as the shortest generation of e_2 only needs the analogy with e_1. We get:

$$U(e_1\&e_2, ego + e_1) = U(e_1\&e_2, e_1) + U(e_1, ego) \quad (5)$$

Equation (5) accounts for the fact that the unexpectedness of coincidences is an increasing function of egocentric closeness, measured by $U(e_1, ego)$. Importantly, this result is obtained without any extensional reasoning, such as taking into account the size of the set to which one implicitly relates when considering alternatives[21].

Complexity and proximity effects

Locations differ in complexity for a given observer. To account for the fact that rare events are more unexpected when they happen in close vicinity, the scope of formula (1) had to be extended to cover continuous domains. Though Kolmogorov complexity is defined only for discrete structures, the complexity of a place is naturally defined as the most concise set of directions that allows finding it. Locating a surface a^2 on a two-dimensional area S requires no more than $\log_2 (S/a^2)$ bits. To generate the location l of a rare event, *e.g.* a building on fire, when one's experience indicates that comparable events occur with spatial density D_e, one needs $C(l) = \log_2 1/(a^2 D_e)$ bits, as $1/D_e$ represents the area of occurrence of one event on average. Determining the same location knowing that the fire occurred at egocentric distance d requires $C(l|d) = \log_2 (2\pi d/a)$ bits. Unexpectedness due to proximity varies as $U(l,d) = \log_2 (1/(2\pi d^2 D_e))$, assuming the complexity $C(d)$ of the distance to the event is $\log_2 (d/a)$ (where $d > a$). If D_e is estimated by $1/(\pi R^2)$ where R is the distance to the closest remembered event,(3) then:

$$U(l,d) = 2\log_2 (R/(d\sqrt{2})) \quad (6)$$

Unexpectedness is found to be an increasing function of rarity (measured by R) and a decreasing function of egocentric distance d.

Proximity effects are also to be observed in coincidences. The report on the double suicide by drowning[23] heavily insisted on the spatio-temporal proximity of the two events, which occurred a few kilometres apart on the same morning. $C(e_2|e_1)$ in formula (2) involves a term $2\log_2 (R/(d_a\sqrt{2}))$, where d_a is the allocentric distance between the two events. The coincidence is correctly predicted to be more unexpected when d_a is small.

Formula (3) predicts the use of landmarks in narrative descriptions. Mentioning the location l of an event may be useless if it is not close enough to produce any

---

[3] It can be shown to be a good estimator – see www.enst.fr/~jld/Data/Closest-occurrence.pdf

unexpectedness. However, if l is close to a well-known landmark such as the Eiffel Tower,[4] then unexpectedness may recover a significant value. The best landmark maximizes $U(l,L+d_L)$, where $d_L > a$ is the allocentric distance from landmark L to the event. From formula (3), we get:

$$L_{opt} = \mathrm{argmin}_L\ (C(L) + 2\log_2 d_L) \qquad (7)$$

We must have $C(L_{opt}) + 2\log_2 d_{L_{opt}} < 2\log_2 d$, otherwise no landmark is used. Landmarks behave like Newtonian attractors with mass $2^{-C(L)}$ on a two-dimensional space. The null landmark (ego) is granted with mass equal to 1 (figure 3).

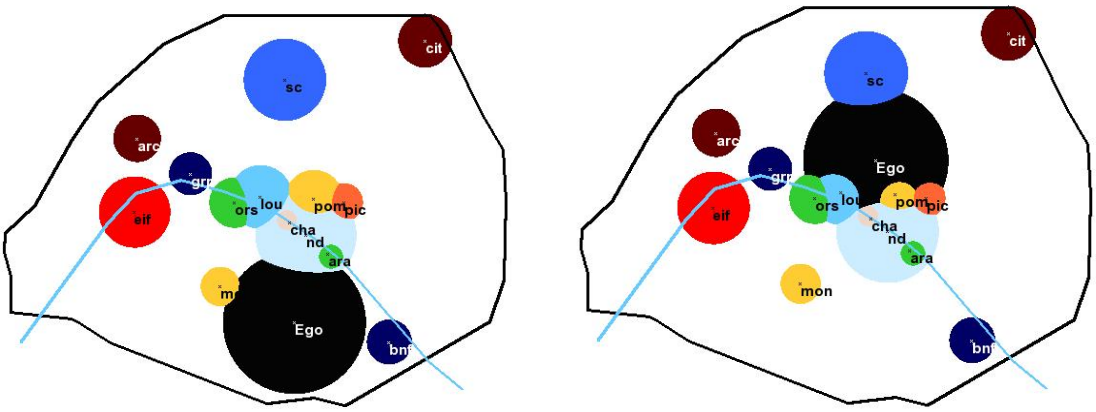

Figure 3: **Landmark influence zones**
These pictures show a map of Paris with the main monuments ranked according to the number of tourist visits. Each monument would be used by a tourist as a landmark within the zone of corresponding colour. The influence of a monument diminishes as 2^{-r} (where r is the rank of the monument in the list) divided by the square of the distance to the point. The largest zone is the *ego* zone, in two successive positions. Outside the coloured zones, no landmark is used except the city of Paris itself.
(an animation can be viewed at: www.enst.fr/~jld/Data/Paris_landmarks.avi)

A similar law holds for time, explaining the importance given to recency in event selection and the frequent use of temporal landmarks for non-recent events[27].

The same phenomenon may explain *topic relatedness*. Each element of the current context of a conversation can be used as a landmark that reduces the complexity of the next topic, making it worth reporting. The mention of F_1 in the current conversation topic makes it free ($C(F_1) = 0$) for the next topic s, which appears more unexpected: $U(s, F_1+F_2) > U(s, F_2)$ as soon as $U(F_2, F_1)$ is non negligible (see equation (4)). A sport car with a towing attachment (F_2) is indeed more unexpected if the previous topic has introduced car accessories (F_1). Abrupt topic change avoidance in casual conversation would not be the manifestation of a pure social convention[28], but would result, at least in part, from a cognitive propensity to maximize unexpectedness.

[4] July 22, 2003, a minor blaze on the upper floor of the Eiffel Tower was reported in French national news media.

The encounter problem

One of the most spectacular problems is the case of a fortuitous encounter s with a person P at location l, as the explicit influence of cognitive complexity is particularly transparent and is not predicted by alternative models. We show that unexpectedness includes the term $C(l) - C(P)$, which explains why surprise is maximal when the location is complex (a small alley lost in an insignificant remote village) and the actual person is simple (*e.g.* a colleague or a celebrity). In what follows, $l(ego)$ and $l(P)$ designate the presence of the witness and P in l. A recursive application of formula (3) gives:

$$U(s, l(ego)+P+l(P)) = C(s) - C(l(ego)) - C(P|l(ego)) - C(l(P)|l(ego)\&P) - C(s|l(ego)\&P\&l(P))$$

$C(l(ego))$ corresponds to the complexity of the plan needed to get to l. It amounts to $C(l)$ in most cases (unless l is difficult to reach). The term $C(P|l(ego))$ equals to $C(P)$, as the complexity of P does not change with *ego*'s position. The term $C(l(P)|\, l(ego)\&P)$ equals zero, as P's position is fully determined using *ego*'s position. We get:

$$U(s, l(ego)+P+l(P)) = C(s) - C(l) - C(P) - C(s \mid l(ego) \ \& \ P\& \ l(P))$$

The difference $C(s) - C(s \mid l(ego) \ \& \ P\& \ l(P))$ corresponds to the standard complexity of generating P and the common presence of *ego* and P in l. If P and *ego* are comparable, if they are colleagues for instance, then the most parsimonious generation of P consists requires $\log_2 N$ bits, where N is the number of colleagues. If $l(ego)$ and $l(P)$ are independent, bringing both protagonists to l requires $2C(l)$ bits. We get eventually: $U(s, l(ego)+P+l(P)) = \log_2 N + C(l) - C(P)$.

If *ego* and P are not compared, for instance if P is a celebrity rather than a colleague, then the standard generation of P may proceed by distinguishing among all individuals that could have been at l, *i.e.* anyone within a range R from l. The value of R must be large enough to include the actual person P, but not larger to keep the generation of s minimal. All these individuals are normally as far from l as P or less. If P is supposed as distant from l as *ego*, the determination of P requires $C(l) + c$ bits, where c is a constant. We obtain finally: $U(s, l(ego)+P) = c + C(l) - C(P)$.

Both computations produce a similar dependence on the complexity of the place and on the simplicity of the person. Though the preceding computations are not trivial, human beings seem to possess specific abilities and heuristics to perform the corresponding complexity estimates spontaneously, since they have a correct perception of all the factors that affect fortuitous encounters.

Note that if the adventure occurred to Q instead of *ego*, then Q must be added to the computation sequence and unexpectedness is diminished by the amount $C(Q)$. Hence the general preference for fist-hand narratives[38], as they are systematically more unexpected.

Cognitive complexity assessment

Cognitive complexity provides a unified and parsimonious framework to account for various factors that have decisive influence on interest: remarkable structure, deviance, and spatial, temporal or social proximity (Figure 4). For a complexity-based model of interest to be predictive, it is essential to have a plausible picture of how human subjects make complexity judgements.

The complexity of structures crucially depends on how the mind analyses them[18]. Their description as a recursive combination of algebraic groups, as in the Generative Theory of Shape[21], suggests that the complexity of a structure is obtained by adding together the complexity of each group entering its description[16]. The validity of this method may be assessed by measuring how reaction times correlate with Gestalt simplicity[29]. Structural unexpectedness can also be estimated by its consequence on probability judgements[16].

The complexity of many situations is assessed by comparison with similar memorised situations, and thus depends on the subject's experience, embedded primarily in prototypes and exemplars[30]. Communication is made possible by the fact that strong between-individuals agreement exists on the prototypes that are formed about familiar objects or about daily events[31]. Actual objects or situations that are close to the centre of the prototype for a multi-attribute resemblance[32] are maximally complex, as discrimination is most difficult there. To assess their complexity, individuals must be able to estimate the numerosity of the corresponding class, *e.g.* by assessing its cognitive 'availability'[15]. Deviant objects or situations are simpler. For distinguishing features with a Gaussian distribution, the complexity of discrimination is expected to decrease as k^2, where k is the number of standard deviations from the mean, until it reaches zero for items that are regarded as unique in their own kind. Human individuals may have an implicit knowledge of this decrease of complexity.

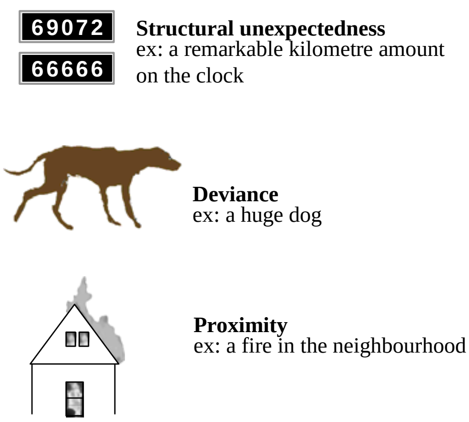

Figure 4: **Sources of unexpectedness**
Simple structures, deviance and proximity are the three main sources of unexpectedness. In each case, the minimal description of the situation is simpler than usual.

Assessing unexpectedness due to proximity in time and space seems straightforward, as it depends on distance estimates as in equation (6). Care should be taken, however, to consider psychological space and time, rather than their physical counterparts. Subjects tend to distort spatial distances[33] and temporal distances[34], for instance by overestimating small amounts and underestimating large ones. Note that subjects spontaneously use spatial[33] and temporal[27] landmarks to locate places and events. A relevant test would be to know whether equation (7) is correct in predicting that the decision to use landmarks depends on their complexity, which would be independently assessed, for instance, through reaction time measures.

Social closeness also has a huge impact on the interest of events, especially in gossip[35,36]. The same anecdote is more interesting if it involves a neighbour rather than that neighbour's cousin. To assess the contribution of social closeness to unexpectedness, one must observe that the number of acquaintances grows exponentially with the number of separating links in the social graph (unless strong geographical constraints bring the increase down to a mere power law[37]). Consequently, the complexity of discriminating among acquaintances grows linearly with social separation. Unexpectedness, and thus gossip interest, will decrease accordingly. The concept of cognitive complexity also accounts for the fact that much gossiping concerns celebrities[35], though famous figures have a bounded social connectedness and are unlikely to be close to all who gossip about them. The reason is that the minimal cognitive determination of celebrities does not go through the graph of acquaintances. Gossiping partners share a list of people they consider prominent. Complexity of prominent figures can be assessed by the logarithm of the rank in that list.

Conclusion

Several aspects of human lives depend on the ability to be surprised at some relevant configurations of the environment. Studies on memory have shown that unexpected and deviant states of affairs are preferentially memorised[39,40,41,42]. Definition (1) offers a precise characterisation of the unexpected and predicts that complexity drop would be a good predictor of memory fixation. It also raises new questions about the function of episodic memory[43].

The notion of unexpectedness offers a new way to found subjective probabilities[16]. An obvious way to link complexity-based and information-theoretic notions of unexpectedness is to define subjective probability by $p = 2^{-U}$, thus considering cognitive complexity as a basic ability and turning probability into a derived notion. With this definition, most situations do not appear within sets of exclusive alternatives and yet can be assigned probability.

Unexpectedness, especially deviance, is strongly connected to newsworthiness[4,6]. The definition of unexpectedness as complexity drop offers a theoretical foundation to several 'news values'[4,44,45] and it indicates how these criteria can be aggregated instead of being merely added[4]. It opens the way to new automated method for personalized news and for ranking streams of news in news search engines[46].

Definition (1) provides a new criterion for narrative relevance (together with other factors like emotion[47]). The prediction is that unexpected situations are preferentially signalled and spontaneously recounted[48], and that the relevance of F to s is measured by

$U(s,F)$. This criterion for narrative relevance has various advantages: it is formal, it offers a gradual measure of interest, it is grounded in universal cognitive abilities and not in convention, and it is open to refutation. It indicates that the content of human conversation may be determined through non-trivial formal computations. Human beings *know* that fortuitous encounters are more interesting when the place is complex and the partner simple. They know that conceptual analogy and physical proximity make coincidences more arresting. They know which details are crucial in a narrative, they know that first-hand narratives are more interesting and they know that abrupt change in conversational topics should be avoided. This knowledge, unexplained by current theories of relevance, appears to follow from one single principle, the maximization of complexity drop.

The predictive power of the notion of unexpectedness suggests that human beings have the fundamental ability to assess the complexity of minimal cognitive descriptions. Kolmogorov complexity has been defined to measure the randomness of mathematical objects. Its inventors may not have anticipated that it captures a basic principle of cognitive activity[19]. Here we showed that complexity drop is a crucial factor involved in the selection, among all experiences, of those which are worth memorizing and telling.

Bibliography

1 Norrick, N. R. *Conversational Narrative: Storytelling in Everyday Talk.* (John Benjamins Publishing Company, Amsterdam, 2000).

2 Maclean, M. & Pinna, L. Distance and news interest: Scarperia, Italy. *Journal. Q.* **35**, 36-48 (1958).

3 Brooker-Gross, S. R. Spatial aspects of newsworthiness. *Geografiska Annaler. Ser. B, Hum. Geogr.* **65**, 1-9 (1983).

4 Galtung, J. & Ruge, M. H. The structure of foreign news: The presentation of the Congo, Cuba and Cyprus crises in four foreign newspapers. *J. Int. Peace Res.* **2**, 64-91 (1965).

5 Shoemaker, P. J., Danielian, L. H. & Brendlinger, N. Deviant acts, risky business and U.S. interests: The newsworthiness of world events. *Journal. Q.* **68**, 781-795 (1991).

6 Pritchard, D. & Hughes, K. D. Patterns of deviance in crime news. *J. Commun.* **47**, 49-67 (1997).

7 Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379-423,623-656 (1948).

8 Weaver, W. Probability, rarity, interest, and surprise. *Sci. Mon.* **67**, 390-392 (1948).

9 Terrell, D. A test of the gambler's fallacy: Evidence from pari-mutuel games. *J. Risk & Uncertain.* **8**, 309 - 317 (1994).

10 Tversky, A. & Kahneman, D. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychol. Rev.* **90**, 293-315 (1983).

11 Kahneman, D. & Tversky, A. On the psychology of prediction. *Psychol. Rev.* **80**, 237-251 (1973).

12 Koehler, J. J. The base rate fallacy reconsidered: descriptive, normative and methodological challenges. *Behav. Brain Sci.* **19**, 1-53 (1996).

13 Nisbett, R. E. & Ross, L. *Human Inference: Strategies and Shortcomings of Social Judgments.* (Prentice Hall, Englewood Cliffs, N.J, 1980).

14 Kahneman, D. & Tversky, A. Subjective probability: A judgement of representativeness. *Cogn. Psychol.* **3**, 430-454 (1972).

15 Tversky, A. & Kahneman, D. Availability: A heuristic for judging frequency and probability. *Cogn. Psychol.* **5**, 207-232 (1973).

16 Dessalles, J-L. A structural model of intuitive probability. In *Proceedings of the seventh International Conference on Cognitive Modeling* (eds Fum, D. Del Missier, F. & Stocco, A.) 86-91 (Edizioni Goliardiche, Trieste, IT, 2006). http://www.enst.fr/~jld/papiers/pap.cogni/Dessalles_06020601.pdf

17 Li, M. & Vitányi, P. *An Introduction to Kolmogorov Complexity and its Applications* (Springer Verlag, New York, 1993).

18 Griffiths, T. L. & Tenenbaum, J. B. Probability, algorithmic complexity, and subjective randomness. In *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (eds Alterman, R. & Kirsh, D.) (Lawrence Erlbaum Assoc., Hillsdale, N. J., 2003). http://web.mit.edu/cocosci/Papers/complex.pdf

19 Chater, N. & Vitányi, P. Simplicity: a unifying principle in cognitive science? *Trends Cogn. Sci.* **7**, 19-22 (2003).

20 Solomonoff, R. J. Complexity-based induction systems: Comparisons and convergence theorems. *IEEE Trans. Inf. Theory* **24**, 422-432 (1978). http://world.std.com/~rjs/solo1.pdf

21 Leyton, M. *A Generative Theory of Shape* (Springer Verlag 2145, New York, 2001).

22 Falk, R. Judgment of coincidences: Mine versus yours. *Am. J. Psychol.* **102**, 477-493 (1989).

23 *Le Figaro* (20 March 2004).

24 Cornuéjols, A. Analogie, principe d'économie et complexité algorithmique. In *Actes des 11èmes Journées Françaises de l'Apprentissage* (Sète, 1996). http://www.lri.fr/~antoine/Papers/JFA96-final-osX.pdf

25 Kern, K. & Brown, K. Using the list of creepy coincidences as an educational opportunity. *Hist. Teach.* **34**, 531-536 (2001). http://links.jstor.org/sici?sici=0018-2745(200108)34%3A4%3C531%3AUTLOCC%3E2.0.CO%3B2-P

26 Falk, R. & Macgregor, D. The surprisingness of coincidences. In *Analysing and Aiding Decision Processes* (eds Svenson, O. & Vári, A.) 489-502 (Adadémiai Kiadó, Budapest, 1983).

27 Shum, M. S. The role of temporal landmarks in autobiographical memory processes. *Psychol. Bull.* **124**, 423-442 (1998).

28 Grice, H. P. Logic and conversation. In *Syntax and Semantics, vol. III, Speech Acts* (eds Cole, P. & Morgan, J. L.) 41-58 (Academic Press, New York, 1975).

29 King, D. L. Gestalts are more closely associated with performance on a discrimination task than are component stimuli. *Am. J. Psychol.* **103**, 37-52 (1990).

30 Medin, D. L. Concepts and Conceptual Structure. *Am. Psychol.* **44**, 1469-1481 (1989).

31 Rosch, E. Principles of Categorization. In *Cognition and Categorization* (eds Rosch, E. & Lloyd, B. B.) 27-48 (Lawrence Erlbaum Assoc., Hillsdale, N. J., 1978). http://originresearch.com/documents/rosch1b.htm

32 Rosch, E. & Mervis, C. B. Family resemblances: Studies in the internal structure of categories. *Cogn. Psychol.* **7**, 573-605 (1975).

33 Tversky, B. Remembering spaces. In *The Oxford Handbook of Memory* (eds Tulving, E. & Craik, F. I. M.) 363-378 (Oxford University Press, New York, 2000). http://www-psych.stanford.edu/~bt/space/papers/rememspacehandbook.doc.pdf

34 Janssen, S. M. J., Chessa, A. G. & Murre, J. M. J. Memory for time: How people date events. *Mem. & Cogn.* **34**, 138-147 (2006). http://www.neuromod.org/publications/papers/MFT/MFT.pdf

35 Levin, J. & Arluke, A. An exploratory analysis of sex differences in gossip. *Sex Roles* **12**, 281-286 (1985).

36 Eder, D. & Enke, J. L. The structure of gossip: Opportunities and constraints on collective expression among adolescents. *Am. Sociol. Rev.* **56**, 494-508 (1991).

37 Csányi, G. & Szendröi, B. Fractal-small-world dichotomy in real-world networks. *Phys. Rev. E* **70** (2004). http://www.csanyi.net/Publications/scaleproof.pdf

38 Coates, J. *Men talk.* (Blackwell, Oxford, UK, 2003).

39 Woll, S. & Graesser, A. Memory discrimination for information typical and atypical of person schemata. *Soc. Cogn.* **1**, 287-310 (1982).

40 Pezdek, K., Whetstone, T., Reynolds, K., Askari, N. & Dougherty, T. Memory for real-world scenes - The role of consistency with schema expectation. *J. Exp. Psychol.: Learn. Mem. & Cogn.* **15**, 587-595 (1989).

41 Stangor, C. & McMillan, D. Memory for expectancy-congruent and expectancy-incongruent information: a review of the social and social developmental literatures. *Psychol. Bull.* **111**, 42-61 (1992).

42 Shapiro, M. A. & Fox, J. R. The role of typical and atypical events in story memory. *Hum. Commun. Res.* **28**, 109-135 (2002).

43 Dessalles, J-L. Storing events to retell them (Commentary on Suddendorf & Corballis: 'The evolution of foresight'). *Behav. Brain Sci.* **30** (2007, in press). http://www.enst.fr/~jld/papiers/pap.evol/Dessalles_07051403.pdf

44 Bell, A. *The Language of the News Media.* (Blackwell, Oxford, UK, 1991).

45 McGregor, J. Restating news values: contemporary criteria for selecting the news. In *Proceedings of the ANZCA 2002 Conference* (ed Power, M.), (Gold Coast, Queensl. 2002).

46 Del Corso, G. M., Gulli, A. & Romani, F. Ranking a stream of news. In *Proceedings of the 14th International Conference on World Wide Web*, 97 – 106 (ACM Press, New York, NY, 2005). http://delivery.acm.org/10.1145/1070000/1060764/p97-delcorso.pdf?key1=1060764&key2=6302340811&coll=GUIDE&dl=GUIDE&CFID=23907380&CFTOKEN=75224440

47 Dessalles, J-L. Le rôle de l'impact émotionnel dans la communication des événements. In *Actes des Journées Francophones 'Modèles formels de l'interaction' (MFI-07)* (eds Lang, J. Lespérance, Y. Sadek D. & Maudet, N.) 113-125 (Annales du LAMSADE, Université Paris Dauphine, Paris, 2007). http://www.enst.fr/~jld/papiers/pap.conv/Dessalles_07022101.pdf

48 van Dijk, T. A. Stories and racism. In *Narrative and Social Control: Critical Perspectives* (ed Mumby, D. K.) 121-142 (Sage Publications, Newbury Park, CA, 1993).
http://www.discourse-in-society.org/teun.html

Dépôt légal : 2007 – 3ème[r] trimestre
Imprimé à l'Ecole Nationale Supérieure des Télécommunications – Paris
ISSN 0751-1345 ENST D (Paris) (France 1983-9999)

Ecole Nationale Supérieure des Télécommunications

Groupe des Ecoles des Télécommunications - membre de ParisTech

46, rue Barrault - 75634 Paris Cedex 13 - Tél. + 33 (0)1 45 81 77 77 - www.enst.fr

Département INFRES