João Ferreira do Amaral FEB 2001

SURPRISE AND UNCERTAINTY

IS IT RATIONAL TO BE RATIONAL?

Introduction

Since the forties that Shannon's theory of information has conceived information mainly as reduction of uncertainty.

Let *T* be a system which has a set of potential states with a value *H* of uncertainty. Then the quantity of information of the observed state of the system is the difference between the amount of uncertainty before and after the observation. That is,

$I = H - H^*$

where H^* is the value of uncertainty after the observation.

If $H^* = 0$, so that the observation sweeps away all the uncertainty we have $I = H$.

As a measure of *H* that is of the uncertainty of the system, Shannon uses Entropy : $H = - \sum p_l \log p_j$ (log base 2).

However, as is easily seen, an information usually means much more than the reduction of uncertainty (and sometimes it even increases it).

Suppose, for instance that two runners A and B start a 5 Km race. We know past performances of both runners so that we give to competitor A 60 % probability of winning and to B 40 %.

Using the concept of Shannon's quantity of information we can say that the message " B has won" has exactly the same quantity of information than the message "A has won" since both reduce uncertainty to *0*.

However, the first message has more meaning because the occurrence it describes is a surprise.

There is then, an element of surprise that is important to consider in the theory of information.

The measure of surprise

In this section we introduce a measure of surprise for a single non repeated information about a system (if it is the case that there are more than one observations the measure of surprise may be for instance the distance between the observed relative frequencies of the states observed and their *a priori* probabilities but this is a case that we don't study here).

Let T be a system with n potential states each with its probability p_j.

Suppose that we make an observation on T from which we get the information that state k has occurred.

Surprise of this information is, by definition

$S_k = log [(max\ p_j) / p_k]$ *(log base 2)*

In the example of the two runners

$S_B = log\ 1.5$

From the definition we get directly:

- For each possible observation there is no surprise when all the states have identical probabilities, unless it occurs a state with probability *0*.

- There is no surprise when the state that occurs is a state with maximum probability
- When $p_k = 0$ that is when the state that occurs has probability *0*, surprise is infinite.

Note that the definition of the states must be done on the basis of relevant features of the events for the observer. To see this, suppose that in the former example the event of A winning was divided in two situations with assumed identical probability: A crossing the line with the left foot or A crossing the line with the right foot.

In this new situation we have three potential states. A winning with the left foot (30% probability); A winning with the right foot (30% probability); B winning (40 % probability).

Of course the event "B wins" would have surprise 0 but only because we considered A divided in two situations that are not relevant.

Surprise and entropy

Looking at the definition of surprise we may say that when there is no previous knowledge about the system there is no surprise.

Conversely, when there is no uncertainty, that is when there is a state with probability *1* and all the others have probability *0*, then surprise is *0* or infinite (in the event of occurring a state of probability *0*).

It is possible to connect more formally surprise and entropy

In our discrete system the state k has the probability p_k of occurring.

Therefore the **expected value of the surprise** (which from now on we call simply **Surprise of the system**) is

$S = \sum p_j S_j$

That is, since $\sum p_j = 1$

$S = \log(\max p_j) - \sum p_j \log p_j = \log(\max p_j) + H$

When the probability of the state of maximum probability is kept constant then surprise and entropy both increase or decrease simultaneously.

However, in most situations the variation of the uncertainty is associated with a variation of the probability of the state with maximum probability, so that in this case S and H may be inversely related.

For instance, if all the states have identical probability, surprise is 0 and the entropy is maximum. If there is an improvement in our knowledge of the system so that the number of states with positive probability is kept constant but with different probabilities, we acquire new knowledge and therefore entropy is reduced and H is lower. However, surprise increases from 0 to a positive value.

It is easily seen that for two states, one with probability p and the other $1-p$ the maximum of surprise is obtained for the value p* such that

$p^* = .782.$

For $.5 \le p \le .782$ S increases and H decreases. For $p > .782$ both S and H decline.

Therefore when knowledge increases surprise may also increase. In social systems this may be an additional cause for instability as will be seen later on.

In the next section we consider surprise in non-conscious systems, that is systems formed by non-human elements.

Surprise and non-conscious systems

The evolution of non-conscious systems as for instance an animal population in a given territory may be studied using the concept of surprise.

Consider the following example.

Suppose an animal population that has two alternative strategies.

Strategy W_1 is a strategy of adaptation and specialization to a given environment. The other is a strategy W_2 of non-specialization.

W_1 will give advantage to the population as long as there are no surprises in the environment . However if the environment is affected by surprising events W_2 may be more advantageous.

Let us give a example.

There are only two possible states for the environment: E_1 with probability p ($p > ½$) and E_2 with probability $1\text{-}p$. These two states correspond to two disjoint events.

Suppose that strategy W_1 (specialization) gives the following results:

The result $R_t = (a+bn)$ is obtained if state E_1 of the environment occurs at time t and $R_t = (c\text{-}dn)$ if E_2 occurs at t, with positive a,b,c e d and n being the number of sequential periods - immediately preceding t - where E_1 was verified.

That means that if there is a sequence of occurrences of E_1 and then one occurrence of E_2 all the gains of specialisation are lost and afterwards a new sequence begins with the next first occurrence of E_1. On the other hand the occurrence of E_2 after a long sequence of E_1 worsens the result.

The strategy W_2 of non specialisation corresponds to the case $b=d=0$, that is for each state the result is always the same. The agent does not learn and adapt itself to the environment.

It is easy to see (see appendix) that following W_1 the expected value at time *0* of the value of *n* at time *t-1* is :

$E_0(n_{t-1}) = (p-p^t) / (1-p)$

For *t* sufficiently large we have

$E_0(n_{t-1}) = p/(1-p) \approx S_2$ that is, the surprise of state E_2.

Therefore, for sufficiently large *t* the expected value at time *0* of the result at time *t* of following W_1 is

$E_0 (R_t) = p(a+ b\, p/1-p) + (1-p)(c- dp/1-p)$

The same for W_2 is

$E_0(R_t) = ap+(1-p)c$

So that W_1 is a better strategy than W_2 if and only if

$b/d > (1-p)/p$.

If we define *log b/d* for positive *b* and *d* as the degree of specialisation of W_1 to the environment when E_1 is the state with the highest probability, this means that W_1 is better than W_2 if the degree of specialisation is larger than the symmetric of the Surprise of E_2 ($S_2=\log p/(1-p)= - \log (1-p)/p$).

That is, the strategy of specialisation is the best under the criterion of the expected value at time *0* of the result at *t* if the surprise of E_2 is large. If this is not the case too much specialised species may eventually be decimated by other less adapted ones.

It is time now to turn to social systems, that is systems where there are self-conscious agents that form expectations about future events and that choose their strategies after assessing expected results.

Surprise and social systems

In a previous paper (Amaral 1998) we considered social systems as systems in which the active information (in the sense of Bohm and Hiley 1993[1]) that directs the system is also subjective information, that is information that has meaning for individuals.

In that paper we used this notion to study an economic system in which prices are simultaneously active and subjective information.

If *Q* is the quantity of active information in the price vector (in the paper we suggest a measure for *Q*, which can also be interpreted as the ability of the active information in guiding the system) and if *H* is the quantity of subjective information (entropy) the ratio Q / H gives a measure of the efficiency of the information that guides the system.

However, if *H* declines the efficiency may not increase.

This can happen if the surprise increases.

An increase in surprise causes a perturbation on the system that may cause a decline in *Q*, that is, in the ability of the active information in guiding the system.

Let us see an example.

Let A and B be two economic agents. A uses an input that is part of the production of B in the previous period. With this input A produces an output, part of which is supplied to B in the same period.

In a formal way *x* being production

$x_{At} = f\,(x_{Bt-1})$

$x_{Bt} = g\,(x_{At})$

with *f* and *g* increasing functions.

[1] According to their definition (p.35) there is active information when a form having very little energy enters into and directs a much greater energy.

We get

$x_{At} = F\ (x_{At-1})$, $x_{Bt} = G\ (x_{Bt-1})$

B has some knowledge of reality and operates with this knowledge. Suppose that there is a stationary state x^* and that in case that there is no surprise B keeps to that stationary state. If he is surprised he leaves the stationary state and the distance increases with the level of surprise.

With this assumption we may write

$| x_{B0} - x_B^*| = z\ (S_0),$ z being an increasing function with $z(0) = 0$ and S_0 the surprise at time 0.

After t periods if $x_{B0} > x_B^*$

$x_{Bt} = G^{(t)} (x_B^* + z\ (S_0))$

and the behaviour of the agents will depend on the behaviour of the respective dynamic systems which may very complex if they are non-linear.

In complex systems the active information may loose or reduce its ability to direct the system and that is why the ratio Q/H may decline even when H declines if this decline in H increases at the same time the surprise.

An other example may be seen in the context of Keynes trade cycle theory.

Keynes and cycles. Thoughts on economic policy

As it is well known Keynes in his *General Theory* advanced some ideas on business cycles in the chapter “Notes on the trade cycle”.

In that chapter he considers that crisis, that is the substitution of a declining trend for an increasing one are some times violent, something that on the contrary does not usually happen in the upswing turn of the cycle.

This fact is explained according to Keynes by the circumstance that in crisis there is a collapse of the marginal efficiency of capital, which, being an expectation is prone to deep variations in the psychological climate.

In other words, surprise has an important role in the theory of the cycle, mainly in what concerns the downswing turn, when a negative surprise change expectations and animal spirits leading to a reduction of the marginal efficiency of capital and therefore to a reduction in investment and because of the multiplier effect in economic activity.

This turn will be sharper the less entrepreneurs are prepared to cope with unexpected situations that is, the less they are able to react positively to surprise.

And this leads us to some thoughts on economic policy.

Uncertainty has a bad name in economics. The more uncertain is the economic environment the less efficient are economic agents. That is why it pays to spend resources in increasing knowledge and therefore in reducing uncertainty.

In economic policy a role is usually given to the authorities in reducing uncertainty (in the sixties indicative planning was justified as an instrument for reducing uncertainty) and most economists think that some kinds of information are public goods that should be supplied by the government.

But is reduction of uncertainty always beneficial?

The truth is that, as we have seen, reduction of uncertainty may be associated with an increase in expected surprise and that means that problems of stability may arise when we reduce uncertainty. Therefore reduction of uncertainty is not always beneficial.

The rational expectations theory assumes that economic agents are rational in the sense that for deciding on their behaviour they use a distribution of probability for the future which coincides with the objective distribution (whatever an objective distribution means).

The justification for this hypothesis (which personally I find preposterous) may perhaps lie in the assumption that non rational agents are less efficient and therefore are expelled by rational agents.

If what we being saying about surprise is true this assumption is not valid.

It does not always pay to be rational, that is to use the "objective" probability distribution for decisions, unless we are prepared to cope with surprise.

The problem is that it is extremely difficult to cope with surprise in a "rational" way since it is not usually possible in advance to assess the consequences of being surprised.

Therefore it is at least controversial to say that the authorities should try to reduce all the types of uncertainty of economic agents. If they do this (assuming that is possible) they are creating conditions for more efficient behaviours but they are perhaps also creating more instability and in the end difficulties for the economic system.

There is I think a good case for keeping a system flexible to cope with surprise even if this mean more uncertainty and less efficiency. In political terms is a good case for tolerance and for not trying to reduce the horizon of variability for groups and institutions.

The more a society is prepared for surprise, that is the more there is variability in the way groups and institutions see society, the more a social system can survive and prosper.

The problem becomes even more complex when we introduce the issue of co-ordination of strategies.

Problems of co-ordination

A co-ordination problem exists when what is more rational for every agent to do is not what is the best for the society as a whole.

Problems of co-ordination may occur due to uncertainty but also because of surprise, as we can see in the following example.

Suppose that there are two agents A_1 and A_2 both with two possible strategies W_1 and W_2 which for a matter of simplification we assume that have exactly the same outcomes for each agent.

There are two possible states E_1 and E_2 with respectively probability p and $1-p$.

The matrices of outcomes for the two states are (remember that we assumed that the outcomes are identical for each agent) where the first digit is the outcome for A_1 and the second for A_2

	State E_1		State E_2	
	A_2 plays		A_2 plays	
A_1 plays	W_1	W_2	W_1	W_2
W_1	(8 8)	(10 3)	(0 0)	(0 2)
W_2	(3 10)	(5 5)	(2 0)	(2 2)

It is easy to see that if $2/7 < p < 2/5$ the best strategy for the society according to the maximum expected value is strategy W_2 but for each player the best strategy is W_1.

Suppose for instance $p = 1/3$.

We have for A_1 (and the same for A_2)[2]

$E(W_1 W_1) = 1/3\ 8 + 2/3\ 0 = 8/3$ and $E(W_1 W_2) = 1/3\ 10 + 2/3\ 0 = 10/3$

$E(W_2 W_1) = 1/3\ 3 + 2/3\ 2 = 5/3$ and $E(W_2 W_2) = 1/3\ 5 + 2/3\ 2 = 9/3$

so that W_1 is really the best strategy for each agent. But of course if both follow W_1 we have $E(W_1 W_1) = 1/3\ 8 + 2/3\ 0 = 8/3$ and if both follow W_2 $E(W_2 W_2) = 1/3\ 5 + 2/3\ 2 = 9/3$ and this is a better outcome.

More generally, suppose that a_{ij} is the outcome for A_1 in state E_1 when he plays strategy W_i and A_2 strategy j and b_{ij} the same for state E_2 assuming both players identical and that $a_{21} < a_{11}$, $a_{22} < a_{12}$, $a_{22} < a_{11}$, $b_{11} < b_{21}$, $b_{12} < b_{22}$ and $b_{11} < b_{22}$.

Then it is easy to see that if

$F_1 = \min [(a_{11} - a_{21})/(b_{21} - b_{11}), (a_{12} - a_{22})/(b_{22} - b_{12})] > (1-p)/p > (a_{11} - a_{22})/(b_{22} - b_{11}) = F_2$

the best strategy for each agent is not the best strategy for the society.

On the other hand if $F_1 > 1 > F_2$ with F_i near 1 this non co-ordination is associated with uncertainty (high entropy). But if $F_1 < 1$ this may be associated not with uncertainty but with a high level of expected surprise. Therefore according to the world that the agents face it may be uncertainty or surprise that is associated with non co-ordination.

---

[2] $E(W_i W_k)$ stands for “the expected result for player A_1 when he follows W_i given that A_2 follows W_k”.

Till this moment we have dealt with surprise supposing that there are no errors of judgement. But of course many surprising situations have in their roots a misconception of the probability of the potential outcomes.

Let us consider this question in the following section.

Surprise and error

Suppose that some agent has a probability distribution for n potential states p_j and suppose that the "objective" probability distribution is $\{q_j\}$.

The expected surprise is therefore

$S = \sum q_j \log (p_k / p_j)$ where $p_k = \max p_j$

For each j this expression can be put in the form

$q_j \log p_k + q_j \log (q_j / p_j) - q_j \log q_j$

summing we get

$S = \log p_k + D (q, p) + H$

where $D(p,q)$ is the Kullback-Leibler mesure of divergence between the two distributions (see, for instance Aoki 1996, p.46 and Joyce, 1999, p.215).

Therefore we can interpret surprise as the sum of three components:

- the maximum probability p_k that the agent considers. It can be used as a measure of the specialisation or adaptation of the agent to the environment. A very specialised agent considers a high value for p_k.
- the error $D(q,p)$ of the agent is assessing the distribution of probability of the states
- the "objective" uncertainty of the environment given by the entropy $H(X)$.

I think that this partition gives a useful framework for analysing systems in a large number of areas including evolution theory and economics.

Conclusion

For a large number of systems there are probably two principles in action.

A principle of *efficiency* that is essential for enhancing individual survival, growth and reproduction

A principle of *surprise* that means that the elements or individuals of the system should be kept open to surprise, which is a principle essential for the survival of the species or the group.

There is a tension between these two principles. The principle of efficiency leads to more adaptation and specialisation but this is not always the best strategy.

As we have seen some models can de thought using these two principles.

Suppose an economy that specialises in a given sector or product. This will increase its competitiveness and growth. But unless this specialisation is compatible with the permanence of less specialised firms, that is with firms that can survive even behaving with less knowledge or always keeping open to alternative environments, the specialisation may be dangerous because is vulnerable to surprise.

In biology some features of the evolution of species may perhaps also be interpreted using this framework.

Economists have traditionally considered uncertainty as a trouble. And in a sense they are right. Dealing with uncertainty means decreasing efficiency. However I am not sure that trying to reduce uncertainty and adapting behaviour to a less uncertain environment is always the best strategy. This paper gives some indication that in some cases this strategy may enhance the performance of the individuals but may endanger the group and finally the individual himself.

Perhaps it is not always rational to be rational.

Appendix

Additional results

Addition

Surprises are additive. That means if there are two independent sets of potential states X and Y we have $S_{XY} = S_X + S_Y$ (this is a direct result of the additivity of the entropies of two independent systems).

That means that when an agent gets more independent information he can also be more surprised.

Maximum surprise

If $\{p_j\}$ is a set of finite potential states, the maximum value for S is obtained when the state of maximum probability has probability p_k given by the equation

$e^{1/pk} (1 - p_k) / p_k = n - 1$

and all the other states have equal probabilities, that is

$p_l = (1 - p\ ) / (n - 1)$ for $l \neq k$

(the proof is straightforward and is obtained from the maximization of S subject to $\sum p_j = 1$)

The expected value $E(n_{t-1})$

At *0* the values of *n* at time *t-1* have the following distribution of probability:

Values of *n*	Probability
0	$1-p$
1	$p(1-p)$
2	$p^2(1-p)$
…	…
$t-3$	$p^{t-3}(1-p)$
$t-2$	$p^{t-2}(1-p)$
$t-1$	p^{t-1}

and the expected value

$$E(n_t)=p(1-p)+2p^2(1-p) + \ldots +(t-2)p^{t-2}(1-p)+(t-1)p^{t-1}=(1-p)[p+2p^2+\ldots+(t-2)p^{t-2}] + (t-1)p^{t-1}=$$

$$= p+p^2+\ldots+p^{t-2} - (t-2)p^{t-1}+(t-1)p^{t-1} = p(1-p^{t-2})/(1-p)+p^{t-1}=(p-p^t)/(1-p)$$

REFERENCES

Amaral, João Ferreira do (1998); *"Complexity and information in economic systems"*, UECE / ISEG 1998

Aoki, Masanao (1996); *"New approaches to macroeconomic modeling"*, Cambridge University Press

Bohm, D;Hiley,B.J.(1993); *"The undivided universe"*. Routledge

Joyce, James M. (1999); *"The foundations of causal decision theory"*, Cambridge University Press

Keynes, John M. (1973); "*The general theory of employment, interest and money"*, Macmillan, Cambridge University Press

Khinchin,A.I.;(1957); *"Mathematical foundations of information theory"*, Dover