

ADAPTIVE INFORMATION AND ANIMAL BEHAVIOUR: WHY MOTORISTS STOP AT RED TRAFFIC LIGHTS.

RONALD K. TEMPLETON* and JAMES FRANKLIN**

*30 Carcoola Road, Saint Ives NSW 2075, AUSTRALIA.

**School of Mathematics, University of New South Wales NSW 2033, AUSTRALIA.

Received 12 December 1991; 2 July 1992

ABSTRACT: To behave in an adaptive way an animal must refer to information. Consider what happens when a motorist comes to traffic lights at red. The motorist usually stops. Stopping avoids the high risk of damage and injury at the cost of one or two minutes. So stopping at red traffic lights is the best way for the motorist to behave. There is a relation between the changing pattern of red, green and amber and the best thing to do. This relation is like an abstract signpost that implies what to do. Traffic lights imply and motorists infer. The information that means something is this 'traffic lights' kind of implication that in an abstract way points to the best way to behave. Information of this kind is adaptive information.

Information varies in accuracy, may be unreliable and is often redundant. So how much is a good thing? The appendix proves mathematically that the more information an animal refers to, the closer to the adaptive way it can behave. Three major predictions follow. First, animals refer to all the information available; second, they ignore variables that have only a random association with adaptive behaviour; and third, they interpret information in an adaptive way. In evolutionary time the animal's characters that perceive and interpret its world could have evolved until the animal now does infer what the information implies, and selection ensures that this happens.

* * *

1. INTRODUCTION

When a newly hatched Herring Gull (*Larus argentatus*) surveys its world, at some moment it sees a red patch near the end of its parent's yellow bill. The chick interprets this red patch as an object to be pecked at and when it does so the parent regurgitates an item of food that the chick eats (Tinbergen & Perdeck 1950). It is obvious enough that obtaining the food is good for the chick: if it did not eat, it would soon die of starvation. Adaptive behaviour can be defined as that which increases an animal's contribution to future generations by enhancing its survival or reproduction. This functional approach to behaviour is called behavioural ecology (Krebs & Davies 1984) or sociobiology (Wilson 1975). Clearly the behaviour of the chick that pecks at the red spot on its parent's bill is functional in this way, and so too is the usual behaviour of all animals. But is behaviour a consequence of function alone?

The red patch on the Herring Gull's bill is not only the place that supplies regurgitated food, but is also an excellent indicator or clue for a good place for the chick to peck. The red patch implies that this is a good place, and that is what the chick infers. The characters of the chick that do the inferring – for instance, the eyes that perceive the red patch and the central nervous system that interprets it in this way – constitute part of the mechanism that the chick uses to regulate its behaviour.

But what is this implication that the red patch has for the chick? Puffy white clouds in blue skies do not have it, nor do red spots of lichen on rocks. Nor does the implication depend at all on whether or not the chick does any inferring. The implication is still there and the chick can refer to it only if its characters have evolved that way. *Implication* means much the same as the everyday use of the term *information*. First, we enlarge on what this everyday information is; then, we show how much of it an animal ought to refer to. Lastly, we show that now one can predict which variables are the ones that animals do refer to.

2. WHAT IS ADAPTIVE INFORMATION?

Infer and imply

It is important to clearly distinguish between *infer* and *imply*. When a motorist comes to traffic lights at red, the motorist usually stops. This is hardly surprising; all motorists know that in the circumstances of red traffic lights, stopping is the best thing to do because one avoids injury, damage and traffic fines at the cost of only one or two minutes. This relation between the traffic lights at red and the best thing for the motorist to do is like an abstract signpost, and it is this relation or abstract signpost that is the

*

*

*

Evolutionary Theory 10: 145-155 (December, 1992).

The editors thank A. Abell and another referee for help in evaluating this paper.

© 1992, The University of Chicago.

implication of the red traffic lights. Motorists usually do infer this and stop. In other words, they usually refer to the traffic lights and then make up their minds to stop. Implication is an abstract property of something that someone could infer when making up his or her mind in a logical way. Inference is the psychological process of making up one's mind. It is an individual's interpretation of an object, or rather a pattern of objects; this interpretation is its estimate of the implication.

For the Herring Gull chick, the red spot on its parent's bill does not incorporate any kind of mechanism that compels the chick to peck at it. It is the chick that decides to peck at the red spot. The characters that perceive the red spot and interpret it in this way are part of the chick. In other words, information is permissive in that its only role is to be available in the same way as a red traffic light is available for a motorist to refer to; the light plays no active part in the decision of the motorist to stop or drive through it.

Traffic lights at red and the implication for the motorist to stop are separate and distinct. Common usage calls traffic lights at red or some other particular pattern of objects simply as 'traffic lights at red' or 'a variable', 'pattern of variables', and so on; one is silent about whether or not any implication occurs. When one wants to include what the traffic lights at red imply for a motorist, or perhaps more cautiously to acknowledge that some kind of implication may occur, then the word 'clue' is appropriate. When one wants to refer only to what the traffic lights at red imply for the motorist, then the term 'item of information' is appropriate. As will be shown later, 'information' is reserved for the implication that occurs from all the available items of information taken as a whole.

So an item of information is an abstract signpost or implication. It is a relation: the relation of the red spot on the yellow bill with the best place to peck, and the relation between the traffic lights at red and the best thing for the motorist to do. Information is not the red spot on the yellow bill itself nor the traffic lights at red. No animal can see an implication, but it can have the kind of eyes that see red spots and it can have propensities in its central nervous system to interpret these red spots in the particular way that infers the best thing to do. It can also have built-in propensities to adjust and improve the way it does so, in other words, propensities to learn only certain things (see Hume 1748 p. 108; Hinde & Stevenson-Hinde 1973).

The best response to a traffic light varies with its colour: when the light is red, it is time to stop; when the light is green, it is time to go. This covariance of patterns persists whether the motorist draws the correct inference or not. With the Herring Gull, the covariance in pattern and response is over space not time. Where the bill is red, it is the place to peck; where the Gull is not red, it is not the place to peck. If the light were stuck on red, or the Gull were red all over, how could redness relate to the best response or to anything else at all? With constant redness there could be no implication. An implication is possible only if two or more alternative patterns, sequences, arrangements or the like occur. And they must in some way vary with the best response. If this covariance is negative, the animal can simply reverse it in interpretation; if it is even more abstruse the animal will require an interpretation that is correspondingly subtle.

On the other hand if the traffic light goes from red to green, but does so at random, or if pecking at the gull's red spot does not alter the probability that the parent will regurgitate food, there is no relation with the best response and so no implication. Clearly if no implication exists, no inference is possible. For information to exist, a pattern of variables must relate to a best response, not necessarily all the time, but at least better than chance. (See Young 1987, pp 6-8, 14-15; Dretske 1983).

Context or circumstances

Early in the breeding season, a Herring Gull will eat an egg placed in its nest (Paludin 1951, in Parsons 1976). As day length changes, the way that it is adaptive for the gull to interpret an egg in its nest also changes. During the ten days or so before it lays its own eggs, the Herring Gull will incubate an egg placed in its nest. Similarly, when a Gannet sees an egg in its nest, it incubates the egg by placing the webs of its feet on top of the egg. But when the egg is pipped, the Gannet places its webs under the egg; otherwise the egg would be crushed and the chick squashed (Nelson 1978). What the clue of the egg implies depends on the other clues present, day length for the Herring Gull and the cracked surface and the sound of the unhatched chick for the Gannet. And in general the behaviour that a clue implies is the adaptive one, crucially depends on that part of the other circumstances or context that consists of clues.

The idea of context has been actively presented by W. John Smith (e.g. 1977) in his studies on communication. However, context is not confined to communication, but is just as important to the way all animals, including these Herring Gull chicks, regulate all behaviour. It follows that the variables that

animals perceive must include all the available clues and that what animals interpret must be that part of what they have perceived that constitutes their information.

A tortuous schedule

Information changes continuously: it is not static. The time of day, day length (season), weather, the abundance and crypsis of prey, the sight of approaching predators and the way companions behave all change. The animal has the complicated task of keeping pace with this cornucopia of clues, each one of which changes according to its own tortuous schedule.

Implication, species and individuals

Do clues imply the same thing from one species to another and from one individual to another? Patterns of ultraviolet light contain important clues for honey bees and patterns of ultrasound contain important clues for some bats. No one has ever shown that any of these variables relates in any way to any human adaptive behaviour. Both Red Deer (*Cervus elaphus*) and Starlings (*Sturnus vulgaris*) refer to day length to regulate their time of fertilization, but Red Deer mate in October (Fletcher 1974) whereas Starlings mate in April (Dawson 1986). One assumes, therefore that Starlings interpret day length in the way that is best for Starlings while Red Deer do so in the way that is best for Red Deer. What is a clue to one animal may be irrelevant to another. And when the same variable is a clue to two animals, even then the significance is likely to vary. In general, the variables that are clues, the way they are clues and the behaviour that is adaptive all vary from species to species, at least in fine detail.

Clues also differ in what they imply from individual to individual. For example a domestic cow learns to identify its own calf within a few minutes after giving birth and cannot thereafter be made to foster a similar calf (Lucretius c. 55 BC). Learning describes the general phenomenon that individuals change the way they respond to particular variables when such changes are adaptive and can only be found from experience (e.g. Thorpe 1963; Shettleworth 1984).

Accuracy, reliability and significance

Clues for flying to cover or for any other behaviour also vary in three other ways: accuracy, reliability and significance. First, accuracy. The clues that relate to the adaptive time for a Herring Gull to breed include day length and dawn temperature. Dawn temperature is only as accurate as a weather forecast, whereas day length is potentially accurate to an astronomical degree. But what kind of day length is available for the Herring Gull? It is one whose accuracy varies with the incidence during the preceding few days and weeks, of cloud cover at dawn and dusk, and also with the incidence of shadow in the place where the gull usually spends dawn and dusk. In general, clues always contain some residual error. Second, clues vary in their reliability. The sun and other celestial clues are very accurate, but clouds in an erratic way may hide them from a bird that at a particular time of day or season has to navigate as best it can. No matter what their accuracy celestial clues are sometimes not available at all. Clues of this intermittent kind are unreliable. Third, clues vary in their significance. A bird that overlooks a single itch to preen, perhaps a largely accurate clue, probably loses little in the way of survival or reproduction, whereas a single missed alarm call to seek cover from a predator may be fatal. And in deciding when to breed a Herring Gull that overlooked day length would be overlooking a clue that is highly accurate and where the consequences, success in rearing young, are important to the animals' contribution to future generations as well. The significance of a clue is the change in the animals' expected contribution to future generations (see Fisher 1958) that would occur if the clue were overlooked or misinterpreted. Significance depends on what job the particular behaviour does, as well as on the accuracy of the relation between the clue and the adaptive way.

Summary

What is adaptive information? It is the general assembly of clues, where a clue is a pattern of objects or events that relates to an adaptive way to behave. The behaviour that a clue relates to crucially depends on the rest of the assembly of available clues. Animals perceive clues, but what they interpret is information. Information changes continuously and the behaviour that it implies varies from species to species and also from individual to individual. Clues vary in three other ways: accuracy, reliability and significance. Clues are permissive and do not themselves determine behaviour or influence it in any way at all. However, selection ensures that an animal's characters evolve until the animal does perceive clues and does interpret the resulting information such that it regulates its behaviour in an adaptive way.

3. HOW MUCH INFORMATION IMPROVES THE IMPLICATION?

On hearing an alarm call a bird flies to cover. Clearly whenever the adaptive way to behave is implied by only one clue, it will benefit the animal to refer to that clue. But sometimes two or more clues have implications for the same adaptive way to behave. For example, when a Herring Gull sees an egg in its nest various features of the egg relate to the behaviour of incubation. These include shape, size, colour and speckled pattern, all of which are clues for incubation. If one of these features changed, then incubation proceeds generally in the same way, but there is likely to be an increase in the time that the gull takes to decide to do so (Baerends & Kruijt 1973). It is as if the information about incubation has become less accurate, and that the delay in settling on the eggs to incubate them is some kind of measure of this change in accuracy. Each of these clues relates to the adaptive way to incubate, so they are redundant. In this case they probably relate to incubation in roughly the same way, so one is not surprised that the Herring Gull refers to all of them to roughly equal degree. What happens when clues are redundant and differ a good deal in the way they relate to the adaptive way to behave?

Both day length and dawn temperature vary throughout the year in a regular way, and a certain pattern of each of these relates to the time of year that is adaptive for a particular species to start breeding. For some passerines that feed on insects it is adaptive in cold years to delay breeding, probably because of the same delay in the seasonal abundance of their prey. However, if one considers a seabird such as the Herring Gull, then the adaptive time to breed is unlikely to be delayed by an unseasonally low pattern of dawn temperature (Lack 1966). So one can assume for the Herring Gull that day length and dawn temperature are redundant clues, and as already mentioned, the variation from year to year in the relation between dawn temperature and the adaptive time to breed, is much greater than that of day length. If a clue of astronomical accuracy were available, referring to another clue could imply no increase in accuracy at all. However, as shown above even day length contains some residual error, and all or almost all information is only available in this kind of way where some increase in accuracy is possible. For a particular clue the residual error often can be expressed as a variance, and the accuracy can then be defined as the reciprocal of this variance. Where the error cannot be expressed as a variance, then accuracy will have to be expressed in a way that corresponds to the kind of error.

Is there a way of combining a clue of poor accuracy (such as dawn temperature) with one of high accuracy (such as available day length), that will improve the estimate based on the high-accuracy clue alone? Where the clues are independent the way is straightforward. It consists of taking the mean of the two clues weighted inversely according to their respective variances (Appendix). Where the clues are partially correlated, the appendix shows that in all cases there is a way, a somewhat complicated one, that gives a better estimate than could be obtained by referring to the high-accuracy clue alone. This means that the accuracy implied by the information available will *always* be improved by one more clue even when that clue is redundant, partially correlated and inaccurate. Does this help to account for the animal's characters that perceive variables and interpret information? And do any predictions follow?

4. HOW DOES PERCEPTION EVOLVE?

How is one to account for the evolution of all or almost all sensory perception? In the beginning, to regulate a particular pattern of behaviour the animal could refer to the first clue that was available. How accurately could the animal regulate this behaviour? It could not do so more than the accuracy of the information itself. In other words the animal's physiology ought to evolve until given that information, the behaviour is as accurate as that of the information. In time other kinds of clues are likely to arise, so how will the animal's physiology evolve in these circumstances? In due course and by chance it will perceive the new clue and interpret the new clue by marginally changing the way it regulates its behaviour. Then selection ensures that this change will continue until it reaches the amount that is adaptive given the correlation and the variance of the new clue and that of the old one. The same reasoning applies to any number of new clues. On the other hand, if one had assumed that animals interpret a redundant clue by taking the simple mean of the two estimates then it is certainly possible to get a result that is worse than would have been obtained by referring to the more accurate clue alone. To illustrate this idea, the appendix shows that an animal that interprets a redundant clue in this way will improve its estimate only when the variance of the redundant clue is less than three times that of the first one. This reasoning probably accounts for the intuitive feeling that poor information can only adulterate good information. But if this were so, the initial interpretation of a second clue would always adulterate the interpretation of the first one and make it less accurate. And to regulate any pattern of behaviour no

animal could refer to more than one clue. The incubation behaviour of the Herring Gull mentioned above and much other behaviour shows that this is not so.

During its life an animal comes across a tortuous schedule of information. For any information, the appendix shows how to weight each clue so as to calculate the best estimate of the adaptive way to behave. Over a number of generations the animal's characters that perceive and interpret this information evolve until the animal does infer what the information implies.

5. WHAT PREDICTIONS FOLLOW?

If one prevents an animal in an experiment from referring to one of its usual clues, what results will follow? First, the animal will still regulate its behaviour about the adaptive way, but the variance will now be larger. And secondly, because of the chance of sampling, in a few instances the animal that refers to fewer clues will approach closer to the adaptive way and not further away. These results will be easier to detect when the accuracy of the clues are similar, but will still occur whatever the respective accuracies. Consider the clues which pigeons use in navigating back to their loft. This question is not fully resolved. For example it is not agreed whether or not pigeons refer to olfactory clues (see Gould 1982). In a recent experiment (Wiltschko, Wiltschko & Jahnelt 1987) homing pigeons were released after treating their nasal cavities with a local anaesthetic so that they would be unable to perceive any olfactory clue. In this experiment the deviation of the mean directions from home in the treated birds increased slightly, but this effect was not completely consistent. Both the increase in deviation and its inconsistency accord in a qualitative way with the hypothesis that these pigeons use both kinds of clues.

Current theory about how animals regulate their behaviour is encompassed by the ideas of causal factors and motivation. By Halliday's definition (1983, p. 106), '...A causal factor is any event, process or change in some condition which can be shown to activate, sustain or inhibit a particular behaviour pattern. Causal factors may be external stimuli, such as the appearance of a predator, the presentation of a mate, or internal states such as the level in blood of a hormone or of glucose. Internal and external causal factors interact to elicit behaviour...' The combined state of all the internal causal factors is termed the animal's motivation. The way an animal responds to an external causal factor (such as the actions of its mate) varies according to its motivation or internal state. Note that a causal factor is defined as something which 'activates...' a particular behaviour much as if it were some kind of agent.

Current theory of behavioural regulation is concerned solely with the animal's characters that do the inferring. It describes what they are, what they respond to, what they do and how they do it. It does not say what is characteristic of the variables that the animal does refer to and what is characteristic of the ones that it ignores. That depends on the idea of adaptive information. By regarding causal factors as if they were agents, current theory effectively ignores the idea of information, and so has no road to the predictions that follow from it.

Information is concerned with the relation between some pattern of objects or events and an adaptive way to behave. Any variable that relates to an adaptive way to behave is a clue for that behaviour, and one could in principle identify all those variables that are clues. Information then implies adaptive behaviour. And one assumes that the animal's characters have evolved until they now infer what the information implies. To do so the animal must refer to all the clues. So one predicts that the animals' characters have evolved until it now does refer to all the clues that it usually comes across (see Hume 1739, p. 170 *et seq*). A second prediction is that the animal will disregard all variables that have only a random association with adaptive behaviour (see Hume 1739, p. 125). A third prediction is that it will interpret these clues in an adaptive way. To test these predictions in navigation or elsewhere means measuring the accuracy and variance of each clue by itself and in conjunction with the other clues available, a procedure that is rarely (if ever) carried out.

Without information to refer to, no animal could regulate its behaviour no matter what characters it had. Nor could any animal regulate its behaviour without the characters that perceive clues and interpret information, no matter how accurate, reliable and significant the information. Information implies while characters infer. Animals need both. Note that behavioural ecology is about the conditions of life that in aggregate determine adaptive behaviour. But how is the animal to know what this is? Information is the sole resource.

6. SOME INFORMATION IS NOT FREE

Information is not always as freely available as sensing day length, temperature or the sound of an alarm call. Sometimes it can be perceived only at a cost. A Barbary Dove (*Streptopelia risoria*) can monitor the approach of a predator at the cost of loss of feeding (Lendrem 1984), and in a contest a Red Deer stag can improve its estimate of its relative fighting ability by sampling one or more exchanges of roars or other displays, all of which incur some cost (Clutton-Brock & Albon 1979). How could one allow for the added complication of cost?

For vigilance, one can do so in a straightforward way: at the adaptive compromise the combined cost of predation and starvation is a minimum. For a long drawn-out contest, the best way to fight is to obtain more information about only two clues, the best estimate, and its accuracy, of one's relative fighting ability, and to do so as cheaply as possible. Usually this is by a sequence of sampling and escalation. Eventually for one contestant the current estimate of relative fighting ability and its accuracy implies that the expected cost until the other contestant gives up is more than the value of the resource. At that point it is adaptive for the first contestant to give up (Enquist 1984; Leimar 1988). When information is not free, then its cost certainly does affect the adaptive way to behave. On some occasions, at least, one can indicate what this effect will be.

7. CONCLUSION

The animal's characters that perceive and interpret the world evolve until the animal infers what the information implies. This gives rise to three predictions. First, animals will refer to all the information; second, they will ignore variables that have only a random association with adaptive behaviour; and third, they will interpret information in an adaptive way.

APPENDIX

HOW POOR INFORMATION IMPROVES GOOD INFORMATION

How should one combine two inaccurate estimates of the same quantity? Let us model the inaccurate estimates by two independent random variables X and Y with the same mean, μ , but different variances σ_1^2 and σ_2^2 . X and Y represent different (inaccurate) clues to the same quantity that is to be estimated. For example, X may be the correct breeding time as estimated by day length, and Y the correct breeding time as estimated by dawn temperature. These estimates have the same mean μ (the correct date), but different accuracies, measured by the variances of X and Y .

Naively, one might try averaging X and Y , that is, taking $\frac{1}{2}(X + Y)$ as the best estimate of the true quantity. However, the average is a more accurate estimate of μ only if X and Y do not differ too much in accuracy — in particular, only if the variance of the less accurate clue (say Y) is less than three times that of the more accurate one. For,

$$\begin{aligned}\text{Var}\left(\frac{1}{2}(X + Y)\right) &= \frac{1}{4}(\text{Var } X + \text{Var } Y) \\ &= \frac{1}{4}\sigma_1^2 + \frac{1}{4}\sigma_2^2,\end{aligned}$$

which is less than σ_1^2 when $\sigma_2^2 < 3\sigma_1^2$.

So, a simple average of two independent clues is not always useful. But there is always *some* weighted average of them that is an improvement over simply choosing the better one, as we now prove. Any weighted average $\lambda X + (1 - \lambda)Y$ of them, for $0 \leq \lambda \leq 1$, will also have mean μ ; is it possible to choose λ so that the variance of $\lambda X + (1 - \lambda)Y$ is less than the variances of both X and Y ?

Indeed it is. Since the variables are independent,

$$\begin{aligned}\text{Var}(\lambda X + (1 - \lambda)Y) &= \text{Var } \lambda X + \text{Var } (1 - \lambda)Y \\ &= \lambda^2 \text{Var } X + (1 - \lambda)^2 \text{Var } Y \\ &= \lambda^2 \sigma_1^2 + (1 - \lambda)^2 \sigma_2^2\end{aligned}$$

The derivative of this with respect to λ is $2\lambda\sigma_1^2 - 2(1 - \lambda)\sigma_2^2$, which we equate to zero to find the minimum. Thus the choice of λ which minimises the variance of $\lambda X + (1 - \lambda)Y$ is:

Adaptive information and animal behaviour

$$\lambda = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

That is, the most accurate estimate of μ occurs when the measurements are weighted in (inverse) proportion to their variance (the more accurate one, with smaller variance, being of course given the greater weight).

The variance of $\lambda X + (1 - \lambda) Y$ for this choice of λ is:

$$\frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

This improvement in accuracy can be appreciable; it is greatest when X and Y have equal variance, in which case $\lambda = 1/2$, that is, one averages them. The variance of the average is then half the variance of X and Y . This means a 30% improvement in accuracy, if accuracy is measured by the standard deviation of the observations. This advantage increases if a weighted average of more than two sources of information is used: the mean of n equally accurate independent observations of the same μ has standard deviation $1/\sqrt{n}$ times the standard deviation of the individual observations.

Note also that in the case of equal variances, *any* combination of the individual observations has a variance less than that of the observations themselves: combination of different sources can be an advantage even if the weightings are not entirely accurate. Further, the variance of the combination $\lambda X + (1 - \lambda) Y$ is nowhere very sensitive to changes in the ratio of the standard deviations of X and Y , σ_2/σ_1 (a change of $x\%$ in σ_2/σ_1 never makes more than a change of $x\%$ in the variance); thus the organism need not know the accuracy of its original estimates very well in order to take advantage of this method.

What of the assumption in the above that X and Y are independent?

It is reasonable in many cases to model inaccurate sources of information by independent random variables. Suppose, for example, a bird is to estimate the correct breeding date by combining its estimates of dawn temperature and day length. It can be expected that dawn temperature varies about its mean for a combination of many meteorological factors, and thus varies approximately normally. On the other hand, inaccuracies in the bird's estimate of day length would arise mainly from the inaccuracy of its internal clock. Although animals' estimates of daylength are very accurate, measurements of a few minutes' change in a day length of 24 hours cannot be perfect — there is a subtle interaction between a "free-running" internal mechanism and clues such as the diurnal pattern of light which reset the clock. Such measurement errors are usually distributed normally, and there seems no reason to suppose any serious dependence between them and variations in temperature. (In humans, time perception is influenced by body temperature, but only slightly: Conroy & Mills, 1970, pp. 108-111). Hence the inaccuracies in two sources of information can typically be presumed independent (and normal, though that is not relevant here), unless there is reason in a particular case to think otherwise.

Departures from independence will be of significance, but will not necessarily prevent the combining of information from being beneficial. Take the case when Y is actually equal to X , and so contains no new information whatever. Then $\lambda X + (1 - \lambda) Y$ is simply X , so its variance is just the variance of X (and Y). Therefore (in this special case) there is no disadvantage to the organism in combining information from these different sources, even if it is mistaken in regarding them as independent.

One can ask in general whether there is ever a cost in combining information from two different sources as if they were independent. That is, if the organism uses the above rule to weight X and Y , but X and Y are in fact correlated instead of independent, can the variance of $\lambda X + (1 - \lambda) Y$ exceed the variance of either X or Y ? In fact it can, although even in the worst case the loss is not very large. The variance of $\lambda X + (1 - \lambda) Y$ in general (not assuming independence), when λ is chosen as above (to minimize variance, when independence is assumed), is:

$$\frac{\sigma_1^2 \sigma_2^2}{(\sigma_1^2 + \sigma_2^2)^2} (\sigma_1^2 + \sigma_2^2 + 2 \text{Corr}(X, Y) \sigma_1 \sigma_2)$$

(using standard results on correlation; see, e.g., Bickel & Doksum, 1977, p. 453). The worst case occurs when X and Y are perfectly correlated ($\text{Corr}(X, Y) = 1$) and the ratio of the standard deviations, σ_2/σ_1 , is $\sqrt{2} + 1$, or about 2.4; in this case the standard deviation of the combination is about 1.2 σ_1 . Thus, even in the somewhat unrealistic worst case, when the organism mistakes perfectly correlated observations for independent ones, combining the observations as if they were independent results in only a 20% loss in accuracy as against using only the better source.

Nevertheless, 20% may be regarded as important, from the point of view of differential survival. Faced with this problem, there are three possible strategies for an organism:

Strategy 1: Combine all uncertain sources of information as if they were independent, by combining them according to the simple weighting rule above.

This strategy will be successful if, but only if, the sources of information being used are almost always independent (or nearly independent). It is conceivable that animals could have adopted this strategy by evolving means of acquiring information that result in independence most of the time (for example, by shielding the mechanisms of time perception from effects of changes in temperature). The strategy remains dangerous if there are in fact highly correlated sources of information in use.

Strategy 2: Find out whether the sources are highly correlated; if so, use only the better one, and if not, weight as above.

How correlated should "highly correlated" be in this context? One needs to know when taking the more accurate source (say X) gives a lower variance than taking the weighted combination with λ chosen as above. The condition can be calculated as:

$$\text{Corr}(X, Y) > \frac{1}{2} (\sigma_1/\sigma_2 + (\sigma_1/\sigma_2)^3)$$

See figure 1:

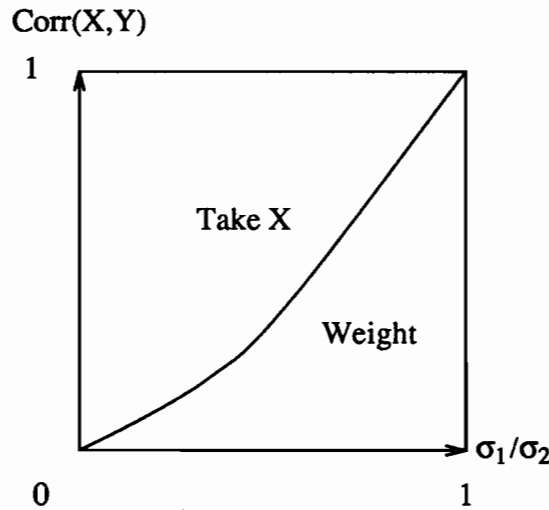


Figure 1
For strategy 2, region where more accurate estimate alone should be taken, versus region where weighted average should be taken

Strategy 3: If the organism knows whether the correlation is high or not, it probably has an estimate of the correlation, which it can use to achieve a better weighting of X and Y than the one that assumes independence. This is the optimal, though most complicated, strategy. One can calculate as above what choice of λ gives the smallest variance for $\lambda X + (1 - \lambda) Y$, where λ may now depend on the correlation. One finds that the best λ is

$$\lambda = \frac{1 - \text{Corr}(X, Y) \sigma_1/\sigma_2}{(\sigma_1/\sigma_2)^2 + 1 - 2 \text{Corr}(X, Y) \sigma_1/\sigma_2}$$

This weighting has a rather different effect in the high-correlation region (meaning the top left region, where the correlation is high compared to the ratio of the accuracies, that is $\text{Corr}(X, Y) > \sigma_1/\sigma_2$), then it has in the low-correlation region, where $\text{Corr}(X, Y) < \sigma_1/\sigma_2$. See figure 2:

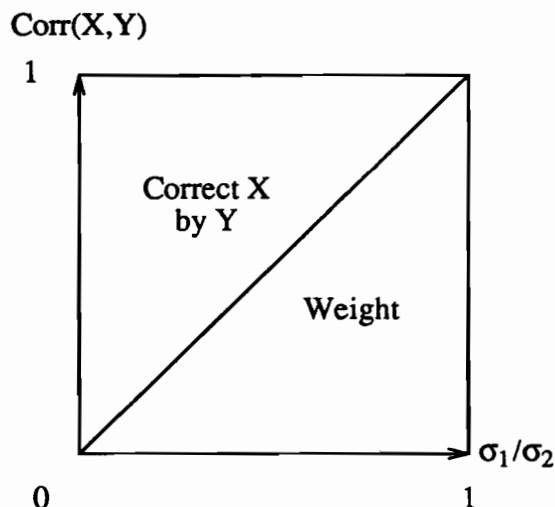


Figure 2
For strategy 3, region where more accurate estimate should be corrected by less accurate, versus region where weighted average should be taken.

In the low-correlation region (bottom right), a weighted average of X and Y is taken, and the result is similar to that of strategy 2, but slightly more accurate. The strategies actually agree along the horizontal axis (where X and Y are independent) and along the right boundary (where $\sigma_1 = \sigma_2$). But at the point $(0.5, 0.3125)$ (which lies on the curve in figure 1), strategy 2 indicates taking X only, giving a variance of σ_1^2 , while the combination given in strategy 3 has a variance of $0.96\sigma_1^2$. This is about a 2% improvement in accuracy, as measured by the standard deviation. This could be considered a worthwhile improvement in some circumstances.

In the high-correlation region, λ is greater than 1, that is, $1 - \lambda$ is negative. This requires a change in interpretation. The expression $\lambda X + (1 - \lambda)Y$ is no longer a "weighted average" of X and Y . Instead, the inaccurate source, Y , is actually given *negative* weight. What this means is that the more accurate measurement, from source X , is *corrected* by the observation Y by going in the *opposite* direction to that suggested by Y . This initially bizarre result can be explained as follows: Suppose an animal is measuring a quantity, say today's dawn temperature, by two methods. X , the accurate one, reads 73° ; Y , the inaccurate one, reads 75° . The animal also knows that X and Y are very highly correlated (so that the readings tend to be both too high or both too low). It should argue that since the inaccurate one is higher, both readings are almost certainly too high; hence it should correct the more accurate one, 73° , by lowering it. In theory, if the correlation were perfect, one could achieve zero variance, or perfect accuracy: if the error in 75° is known to be twice the error in 73° , then the best estimate must be 71° . Such high correlations are of course not to be expected in the wild. Nevertheless, negative weightings of inaccurate sources are an unexpected prediction of the theory that all sources of information are referred to.

The calculations required by this model are well within the capabilities of the brain of higher animals. The human brain has been observed to be an effective "intuitive statistician", in combining information in subconscious ways which are presumably independent of specifically human abilities like language (Peterson & Beach 1967; Kelley, 1973). This is to be expected from the brain's neural network architecture, which is perfectly adapted to computing weighted averages of inputs, with the weights being tunable by evolution and individual experience. (Smolensky 1986, Hinton & Anderson 1989, ch. 1). Also, contrary to appearances, calculating a correlation is very easy, as a correlation is simply the degree of pattern match between two series of observations. That is, the organism only needs to notice if two qualities vary together. Of course organisms do this accurately all the time in inductive reasoning and prediction, which involve first correlating past co-occurrences of a phenomenon and a sign of it. Various classical experiments on conditioning do indeed confirm that rats are very accurate in their estimates of covariation (summary in Holland *et al.* 1986, section 5.2.1). It follows that experiments with rats could be devised to test which, if any, of the strategies just outlined are followed,

but we leave doing so to those more skilled in experimental design. Experiments with robots that rely on many sensors will also cast light on good strategies for combining uncertain information; investigations have begun, but are in the early stages (Luo, 1990).

From the evolutionary point of view, once the initial investment of a powerful brain has been made, the cost of information processing in combining information as outlined here is small. It would certainly be incredible if animals' characters had not evolved to take advantage of these simple means of increasing the accuracy of information.

ACKNOWLEDGEMENTS

Alison Abell redrafted most of the paper, which clarified the argument and greatly improved its readability. Editorial changes by Leigh van Valen likewise have improved it. Comments by David Croft, Leslie Hutley and Maurice Pepper helped to improve earlier drafts. Gwenda Cusack typed numerous drafts. It is a pleasure to thank them.

REFERENCES

- Baerends, G.P. & Kruijt, J.P. 1973. Stimulus selection. In: Hinde, R.A. & Stevenson-Hinde, J. (Eds), *Constraints on Learning: Limitations and Predispositions*, pp. 23-50. London: Academic Press.
- Bickel, P.J. & Doksum, K.A. 1977. *Mathematical Statistics*. San Francisco: Holden-Day.
- Clutton-Brock, T.H. & Albon, S.D. 1979. The roaring of Red Deer and the evolution of honest advertisement, *Behaviour*, 69, 135-144.
- Conroy, R.T.W.L. & Mills, J.N. 1970. *Human Circadian Rhythms*. London: Churchill.
- Dawson, A. 1986. The effect of restricting the daily period of food availability on testicular growth of Starlings, *Sturnus vulgaris*, *Ibis*, 128, 572-575.
- Dretske, F.I. 1983. Précis of *Knowledge and the Flow of Information*, *Behavioural and Brain Sciences*, 6, 55-90.
- Enquist, M. 1984. *Game Theory Studies on Aggressive Behaviour*. Stockholm: University of Stockholm.
- Fisher, R.A. 1958. *The Genetical Theory of Natural Selection*. New York: Dover.
- Fletcher, T.J. 1974. The timing of reproduction in the Red Deer (*Cervus elaphus*) in relation to latitude, *J. Zool.*, 172, 363-367.
- Gould, J.L. 1982. The map sense of pigeons, *Nature*, 296, 205-211.
- Halliday, T.R. 1983. Motivation. In: *Animal Behaviour. 1. Causes and Effects*. (Ed. by T.R. Halliday & P.J.B. Slater), pp. 100-133. Oxford: Blackwell Scientific Publications.
- Hinde, R.A. & Stevenson-Hinde, J. (Eds) 1973. *Constraints on Learning: Limitations and Predispositions*. London: Academic Press.
- Hinton, G.E. & Anderson, J.A., eds, 1989. *Parallel Models of Associative Memory*. Hillsdale, NJ: Erlbaum.
- Holland, J.H., Holyoak, K.J., Nisbett, R.E. and Thagard, P.R., 1986. *Induction: Processes of Inference, Learning, and Discovery*. Cambridge, Mass: MIT Press.
- Hume, D. 1739. *A Treatise of Human Nature* (Ed. by L.A. Selby-Bigge & P.H. Nidditch 1978). Oxford: Oxford University Press.
- Hume, D. 1748. *An Enquiry Concerning Human Understanding* (Ed. by L.A. Selby-Bigge & P.H. Nidditch 1975). Oxford: Oxford University Press.
- Kelley, H.H. 1973. The process of causal attribution, *American Psychologist*, 28, 107-128.
- Krebs, J.R. & Davies, N.B. 1984. *Behavioural Ecology: An Evolutionary Approach*. 2nd ed, Oxford: Blackwell.
- Krebs, J.R. & Dawkins, R. 1984. Animal signals: mind-reading and manipulation. In: Krebs & Davies 1984, pp. 380-402.
- Lack, D. 1966. *Population Studies of Birds*. Oxford: Clarendon.
- Leimar, O. 1988. *Evolutionary Analysis of Animal Fighting*. Stockholm: Stockholm University Press.
- Lendrem, D. 1984. Flocking, feeding and predation risk: absolute and instantaneous feeding rates, *Anim. Behav.*, 32, 298-299.
- Lucretius C55 BC. *On the Nature of Things*. Trans R.E. Latham, 1951. Harmondsworth: Penguin.
- Luo, R.C. ed. 1990. *Journal of Robotic Systems* 7(3), Special issue on multisensor integration and fusion for intelligent robots.

- Parsons, J. 1976. Factors determining the number and size of eggs laid by Herring Gulls, *Condor*, 78, 481-492.
- Peterson, C.R. & Beach, L.R. 1967. Man as an intuitive statistician, *Psychological Bull.*, 68, 29-46.
- Shettleworth, S.J. 1984. Learning and Behavioural Ecology. In: Krebs and Davies 1984, pp. 170-194.
- Smith, W.J. 1977. *The Behaviour of Communicating*. Cambridge, Mass: Harvard University Press.
- Smolensky, P. 1986. Neural and conceptual interpretation of PDP models. In: *Parallel Distributed Processing* (Ed. by J.L. McClelland and D.E. Rumelhart) vol. 2, ch. 22. Cambridge, Mass: MIT Press.
- Thorpe, W.H. 1963. *Learning and Instinct in Animals*. London: Methuen.
- Tinbergen, N. & Perdeck, A.C. 1950. On the stimulus situation releasing the begging response in the newly hatched Herring Gull chick, (*Larus argentatus argentatus* Pont.), *Behaviour*, 3, 1-39.
- Wilson, E.O. 1975. *Sociobiology: The New Synthesis*. Cambridge, Mass: Belknap Press.
- Wiltschko, W., Wiltschko, R. & Jahnel, M. 1987. The orientation behaviour of anosmic pigeons in Frankfurt a.M., Germany, *Anim. Behav.*, 35, 1324-1333.
- Young F.H. 1987. *The Nature of Information*. New York: Wiley.