Prospects for probabilistic theories of natural information

Ulrich E. Stegmann, University of Aberdeen

Abstract

Much recent work on natural information has focused on probabilistic theories, which construe natural information as a matter of probabilistic relations between events or states. This paper assesses three variants of probabilistic theories (due to Millikan, Shea, and Scarantino and Piccinini). I distinguish between probabilistic theories as (1) attempts to reveal why probabilistic relations are important for human and non-human animals and as (2) explications of the information concept(s) employed in the sciences. I argue that the strength of probabilistic theories lies in the first project. Probabilityraising can enable organisms to draw specific inferences they otherwise could not entertain and I show how exactly they help to explain the behaviour of organisms. In addition, probability-raising warrants inferences by providing incremental inductive support.

Introduction

We tend to think of some events as carrying information: the ring of the doorbell carries the information that someone is at the door; footprints in the snow mean that someone walked by; smoke on the horizon carries the information that there is a fire. In his book *Knowledge and the Flow of Information,* Dretske (1981) proposed the most influential theory of what became known as 'natural information'. At its heart was the idea that natural information is a mind-independent, lawful relation between states or events in the world, e.g. between fire and smoke. This relation can be expressed in probabilistic terms: smoke carries the information that there is fire just in case the probability of fire given the occurrence of smoke is one, and less than one otherwise.[1]

Many commentators have criticised as too strong Dretske's requirement that signs must raise probabilities to one (e.g., Suppes, 1983, Godfrey-Smith, 1992, Millikan, 2000, Cohen and Meskin, 2006, Scarantino and Piccinini, 2010). An event rarely, if ever, increases to one the probability of another. Many events or states do not *always* co-occur with others, but only more or less frequently (as when 60% of *A*s co-occur with *B*s), and yet we seem to be able to learn from such imperfect statistical associations. This intuition motivates the theories of information defended by Millikan (2000, 2004), Shea (2007), Scarantino and Piccinini (Piccinini and Scarantino, 2010, Scarantino and Piccinini, 2010, 2011), and Skyrms (2010).[2] Probabilistic theories, as we might call them, maintain that imperfectly related events *can* carry natural information, and they then aim to specify the conditions under which they *do*. It is with respect to these further conditions that the authors' positions diverge. The several notions of natural information pick out distinct probabilistic relations, i.e. degrees of coincidence or probability-raising (the difference will be explained below).

This paper starts by distinguishing between two related but different philosophical projects about natural information. Both are valuable, but it is useful to be clear about what is at stake in each (section 1). The following section (2) clarifies the difference between degrees of coincidence and probability-raising. I then briefly describe and compare three variants of probabilistic theories, due to Ruth Millikan, Nicholas Shea, and Andrea Scarantino and Gualtiero Piccinini, respectively[3] (section 3). Section 4 focuses on the possible significance of probability-raising for organisms. I will argue that probability-raising can enable organisms to infer certain events from others (section 4.1) and that it can provide externalist justification for the inferences (section 4.2). The last section (5) assesses the prospects of using probabilistic theories in order to explicate the information concept(s) used in the behavioural and cognitive sciences.

1. Two projects about natural information

Proponents of probabilistic theories tend to pursue simultaneously two closely related but distinct projects. It will be useful to distinguish them from the outset.[4]

One project is about characterising certain probabilistic relations, like probability-raising, and exploring their significance for the lives of organisms. Advocates of probabilistic information (implicitly) engage in this sort of project when they claim that probabilistic relations allow animals and humans to *infer* (e.g. Scarantino & Piccinini, 2010, p. 318) or come to *know* one event from the other (e.g. Millikan, 2004, p. 37). In the course of such a project one may also explore probabilistic relations from the point of view of naturalized epistemology (e.g. Milllikan, 2004, p. 39-40) and ask whether they *warrant* inferences (more on this below). The probabilistic relations that enable these contributions might be called 'natural information' or 'probabilistic information', perhaps because they capture some of our intuitions about information. But for an enquiry into the significance of probabilistic relations for organisms this is a terminological question and of secondary interest. For a project of this kind, probabilistic theories of information are primarily theories about how, and to what ends, human and non-human animals can employ the probabilistic relations that exist in the world.

Another project is about explicating the concept(s) of information employed within certain domains of interest. A natural starting point is to observe that laypeople and scientists use the term 'information' in various contexts. The project then explores whether 'information' picks out something specific in the world, what features that phenomenon has, whether it changes from context to context, and whether it can do the things the information concept is supposed to do in a given domain of application. Proponents of probabilistic information rely on an explicatory project of this kind when they propose, as a desideratum, that theories of natural information ought to capture the notion of information at work in a discipline like cognitive science. For instance, Scarantino and Piccinini (2010, p. 328) reject the veridicality thesis (roughly, that *A*'s carrying natural information about *B* implies *B*) because it "stands in the way of our understanding of the role played by information in the descriptive and explanatory efforts of cognitive scientists and computer scientists"; veridical notions of information, if appropriate at all, "should not be understood as germane to the main notions of information used in cognitive science and computer science." For an exploratory project, a theory of information is a primarily a theory about the precise content of information as this concept is used in the sciences or everyday life, and a theory about what it picks out. A *probabilistic* theory of information would then be the view that the concept picks out a probabilistic relation.

The extent to which the two projects converge can be left open to investigation. It may turn out that 'information' in the sciences picks out probabilistic relations *and* that organisms can employ probabilistic relations for many things. Or it may turn out that while the latter is true, the term 'information' does not pick out probabilistic relations, and so on.

Keeping the two approaches distinct helps clarifying what is at stake. The bulk of this paper takes the first line, but I will have to say something about the second approach, as well.

2. Degrees of coincidence and probability-raising

The term 'correlation' is often used ambiguously, sometimes referring to degrees of coincidence but more frequently to probability-raising. The degree to which A coincides ('correlates') with B is the degree to which As co-occur with Bs. If As co-occur with Bs 20% of the time, then A's degree of coincidence with B is 20% or, put in probabilistic terms, the conditional probability of B given A is $p(B|A) = 0.2$.[5]

Alternatively, A can 'correlate' with B in the sense of raising B's probability. Suppose Bs never occur without As and As' degree of coincidence with Bs is 20%, as before. The conditional probabilities of B given not-A is then $p(B|¬A) = 0$, but given A it is $p(B|A) = 0.2$. That is, A's occurrence increases B's probability by 20%.

The degree of coincidence is distinct from probability-raising. This can be seen by the fact that A's having a certain degree of coincidence with B does not imply that A raises B's probability by the same or by any degree. We only need to assume that, as above, A's degree of coincidence with B is 20%; but we now also assume that Bs sometimes occur in the absence of As, say 20% of the time. In this case the two conditional probabilities are equal, i.e. $p(B|¬A) = p(B|A) = 0.2$. So despite the positive degree of coincidence, A's occurrence does not increase B's probability.

3. Three probabilistic theories of information

An early articulation of probabilistic information is found in Millikan's work (2000; 2004). Millikan called this type of natural information "softer natural information", "informationC" ('C' for correlation, Millikan, 2000) or "local information" (Millikan, 2004). Although Millikan gradually modified and eventually distanced herself from the idea of local information, due to problems with reference classes and the exclusion of singular events (Millikan, 2007, 2013), it remains an important version of probabilistic information. Millikan's basic idea was that "correlations" enable us to learn one thing from another even when the correlations are imperfect (see below for her notion of correlation):

"Natural signs bearing information$_C$ are, as such, instances of types that are correlated with what they sign, there being a reason, grounded in natural necessity, why this correlation extends through a period of time or from one part of a locale to another." (Millikan, 2000, p. 124).

Bearers of local information are tokens of types of events or states of affairs (As) that correlate with some other types (Bs). A condition that Millikan imposes explicitly on local information is that the correlation between As and Bs persist "for good reason" (Millikan, 2000, p. 121) and not merely by accident (Millikan, 2000, p. 124; 2004, p. 44). Correlations are non-accidental, for Millikan, if there is a direct or indirect causal relation between As and Bs, or if two causally independent trajectories of events run in parallel. An example of the latter is the correlation between the direction of geomagnetic north and the presence or absence of oxygen-poor ocean water in the northern hemisphere, a correlation some marine bacteria exploit for orientation (Millikan, 2004, p. 44). Non-accidental correlations may be restricted to certain spatio-temporal domains, rather than hold universally, and therefore users must remain in the domain in which the correlation holds in order to employ them successfully (Millikan, 2004, pp. 42-43).

A further condition concerns the strength of correlation required for local information: "If the question arises how strong the nonaccidental correlation between As and Bs must be within a domain for the As that do correspond to Bs to count as locally recurrent signs, the following is what is important. A strong enough correlation to count in determining a local sign to be such is one that is strong enough *to have actually influenced* sign use, either through genetic selection or through learning." (Millikan, 2004, p. 44, my emphasis).

It is tempting to understand this condition simply as the requirement that the correlation must be strong enough so as to be usable by organisms.[6] But note the use of the past tense. As formulated, local signs require correlations that are strong enough to *have already influenced* sign use, not strong enough to potentially influence sign use. On this reading, local information and natural signs are user-dependent. For the extent to which a given correlation actually influenced sign use depends not only on the existence of the correlation, but also on an organism having experienced the requisite evolutionary or learning history with respect to that correlation. Local information is then better construed as a three-place relation between As, Bs, and some user. This would be a significant departure from the way natural information is usually understood. And it would be somewhat misleading to assert that organisms "tap into channels of softer natural information" (Millikan, 2000) because there is no information 'out there' for organisms to tap into; out there are just correlations. These correlations *become* natural information for those organisms that manage to exploit them. The

three-place interpretation chimes well with Millikan's more recent views. Millikan's (2004, 2007) persistent concern with determining the reference classes for correlations has led her to give up the entrenched view that natural information is an objective commodity out there in the world. She now argues that natural information is an affordance (Millikan, 2013).[7]

Millikan speaks of "correlations" without elaborating which kind of probabilistic relation is intended. On the one hand, when discussing the non-accidental correlations required for local information, Millikan is concerned with their *strength* only (cashed out in terms of the influence on sign use), as though not concerned with probability-raising. She also insists that her notion of local information is distinct from Shea's correlational information (Millikan, 2007, p. 446), whose central feature is probability-raising. On the other hand, Shea (2007, p. 420) takes his probability-raising account as explicating Millikan's local information. Furthermore, although Millikan (2007) rejects Shea's construal of the relation between his and her account, her rejection is motivated by issues other than probability-raising. Finally, her examples of natural signs (e.g. bee dances and beavers' tail-splashing) do not settle the issue because they are instances of both probability-raising and levels of coincidence. Be that as it may, it is worth noting that adding probability-raising as an explicit condition of local information would be in the spirit of Millikan's general approach. For local information is intended as "at root an epistemic notion" (Millikan, 2004, p. 37), by which she means that when A is a local sign for B then we can learn something about B's presence from A's occurrence. If learning requires probability-raising, as research on animal learning suggests (see below), then refining local information accordingly would be in line with an epistemic notion of natural information.

For Shea (2007), levels of coincidence are not critical for natural information. What matters instead is whether one event makes another more likely:

> "R carries the correlational information that condition C obtains iff for a common natural reason within some spatio-temporal domain D: chance (C | R is tokened) > chance (C | R is not tokened)" (Shea, 2007, p. 421)

On this account even a strong degree of coincidence of, say 70%, may be insufficient for R to carry the "correlational information" that C obtains. If C occured as often in R's absence as it did in R's presence [$p(C|R) = p(C|not\text{-}R) = 0.7$], then R would *not raise* C's probability and therefore would not carry the information that C obtains. By contrast, if it were only the level of coincidence that mattered, a 70% correlation could be strong enough for R to carry natural information.

The primary relata of the informational relation (R and C) for Shea are *types* of events, not tokens. This is clear from the sentence preceding the quote above, in which R is referred to as the "non-semantically-individuated *type* R" (p. 421, my emphasis). But elsewhere in his paper, Shea

appears to allow tokens to carry correlational information as well.[8] The idea seems to be that, in the first instance, one type of event carries information about the other in virtue of one making the other more likely, and that tokens carry information in virtue of their types carrying information. Like Millikan, Shea (2007) excludes accidental correlations from qualifying as informational relations. He demands that a correlation must be sustained by a "common underlying reason" (p. 421), e.g. a causal relation. Shea motivates the exclusion of accidental correlations by arguing that they are explanatorily empty.

Two further features of Shea's (2007) account are worth noting. First, Shea construes probabilities as objective and "not merely a matter of frequencies". Second, non-accidental correlations need not have influenced sign use in order to qualify as correlational information. Consequently, R carries information about *any* state whose probability is increased by R's occurrence (p. 420).[9]

The fundamental idea of Scarantino and Piccinini's (2010) "probability raising theory of natural information" is very similar to Shea's (2007): a state of affairs (*s* is *F*) carries information about another (*o* is *G*) just in case the first increases the probability of the latter:

"(PRT$_N$) If a signal *s* being *F* carries natural information about an object *o* being *G*, then P(*o* is *G*|*s* is *F*)>P(*o* is *G*|(~*s* is *F*))." (Scarantino and Piccinini, 2010, p. 317)

Although probability-raising is only a necessary condition for carrying information according to (PRT$_N$), it is clearly intended as a sufficient condition, as well.[10]

The primary relata of the information relation in Scarantino and Piccinini's account are *tokens*, not types (unlike in Shea's account). The two relata ('*s* is F' and '*o* is G') are construed as Kim-events, i.e. as property instantiations at a time (Scarantino and Piccinini, 2010, p. 316). So interpreted, (PRT$_N$) not only implies that the *information* relation holds between tokens, but also that the *probabilistic* relation holds between tokens. This raises the question of whether the token-level probabilistic relation is derived from a probabilistic relation at type-level or whether it is underived (as would be the case on some propensity views of probability). The following remark suggests the former: "On the view we propose, an event token *a* of type *A* carries natural information about an event token *b* of type *B* just in case *A reliably correlates* with *B*" (Piccinini and Scarantino, 2011, p. 22). In this remark the information carriers are still tokens, but the tokens carry information in virtue of their *types* correlating with one another. This seems to imply that tokens can carry information without the tokens bearing an underived probabilistic relation to one another. I will return to this point.

Unlike Millikan (2004) and Shea (2007), Scarantino and Piccinini *allow* correlations to be accidental (as can be seen from both [PRT$_N$] and informal articulations of their account).[11] Another unique feature is that (PRT$_N$) refers to information "*about* an object *o* being *C*" (my emphasis). Scarantino and Piccinini (2010, p. 316) stress that probabilistic information is not information that *o is C* but, rather, that *o* is *probably C*. It is not only the relation between sign and signified that is probabilistic, as in Millikan (2004) and Shea (2007), but the signified as well. While it remains open how Scarantino and Piccinini's (2010) move from one idea to the other, I will argue below that it provides them with the resources to deal effectively with 'wild' tokens.

Definition (PRT$_N$) offers a concise and formal statement of Scarantino and Piccinini's (2010) account. The statement does not, however, include the notion of reliability, which appears in several other places, including informal summaries of their view such as the following: "the core idea of a probabilistic theory of information is that signals carry natural information about anything they *reliably* correlate with" (Scarantino and Piccinini, 2010, p. 318, my emphasis, see also Piccinini and Scarantino, 2011, as quoted above). Reliability normally indicates a degree of strength. When *A* and *X* raise *B*'s probability by 60% and 20%, respectively, then *A* is the more reliable sign of *B*. Along these lines, Piccinini and Scarantino sometimes suggest that more reliable signs carry more information than less reliable ones (e.g. 2011, p. 22). However, Scarantino and Piccinini appear to give reliability a different meaning when they say that "reliable correlations are the sorts of correlations information users can count on to hold in some range of future and counterfactual circumstances" (Scarantino and Piccinini, 2010, p. 318). This sentence suggests that probabilistic relations are more reliable when they hold over a larger range of future and counterfactual circumstances. Reliability thus appears to concern the range of counterfactual circumstances and time frames across which a probabilistic relation of a given strength obtains. An example may help to see the difference. Suppose that footprints of a certain shape on the forest floor dramatically increase the probability of quails living in the forest (e.g. by 94%). However, if the same forest sustained a large population of pheasants that left indistinguishably similar footprints, then the increase in probability would be smaller (e.g. by 54%). So the probabilistic relation in the actual forest is very reliable in the ordinary sense of being strong, but not reliable in the sense of extending into the counterfactual, pheasant-inhabited forest.[12] The informal characterisation thus suggests that (PRT$_N$) is incomplete. A more complete statement of Scarantino and Piccinini's account would include the claim that a token carries natural information about another token to the extent that (1) one increases the probability of the other [as per (PRT$_N$)] *and* (2) the probability increase holds across a large range of counterfactual circumstances. Since (2) is independent of (1), (2) represents an additional condition on natural information.[13]

Both Shea and Scarantino and Piccinini tie natural information to probability-*raising*. However, Scarantino accepts that a "signal *s* being *F* can also carry (negative) natural information about *o* being *G* by lowering the probability that *o* is *G*" (Scarantino & Piccinini, 2010, footnote 7, p. 317). So presumably, *s*'s being *F* carries the information that *o* is *not G* with some degree of probability if p(*o* is *G*|*s* is *F*) < p(*o* is *G*). Skyrms (2010) has gone farthest in this respect by arguing that an event or state of affairs carries natural information about another when the former *changes* (raises *or* lowers) the probability of the latter.[14] However, in what follows I will continue referring to probability-raising because of its central place in the work of Shea and Scarantino & Piccinini, and because the arguments presented here generalise to probability-changing.

It is instructive to compare the three accounts above with respect to a challenge that arises specifically for probabilistic theories of natural information. Probabilistic theories maintain that natural information can be sustained by imperfect probabilistic relations. A probabilistic relation between *A*s and *B*s is imperfect if some *A*-tokens occur without *B*-tokens. Do these 'wild' *A*-tokens still carry the information that there are *B*-tokens? Note that this question is about how probabilistic theories deal with wild tokens; it does not concern our intuitions about them.

Shea (2007) does not answer the question directly. But he accepts that "even instances [of R] for which C does not obtain are instances of the type which carries correlational information" (Shea, 2007, p. 420). If tokens can carry correlational information in virtue of belonging to a type that carries it, then this remark implies that wild R-tokens carry the information that C-tokens obtain. Such information would be false: an R-token would carry the information that a C-token obtains in the absence of a C-token. Kraemer (2013) argues that allowing natural information to be false is a major problem for Shea's (2007) account because it flies in the face of the widely shared assumption that natural information cannot be false, an assumption accepted by other proponents of probabilistic information, like Millikan (2004) and Scarantino and Piccinini (2010). It is also worth noting that natural information is usually distinguished from representational content by holding that only representations allow for errors (e.g. Millikan, 2004, p. 31). Accepting the possibility of misinformation (i.e. false natural information) would therefore demand a new way of drawing this distinction.

Millikan (2000, p. 124, 2004, p. 44) insists that only those *A*s carry information about *B*s that connect in the "usual" way with *B*s. That is, if *A*s are usually caused by *B*s, then an *A*-token not caused by a *B*-token does not carry the information that *B* obtains. While Millikan's response saves natural information from the possibility of falsity, and thus maintains the traditional distinction between representation and natural information, it also undermines the very idea that imperfect probabilistic relations may suffice for natural information. Natural signs are tokens of a type that

"correlates" for a reason with some state of affairs *B* (Millikan, 2000, p. 124). Now suppose most *A*-tokens are caused by *B*-tokens, but some are not; then both subsets of *A*-tokens are instances of the imperfectly "correlated" type *A*. If *A*-tokens carry information in virtue of being instances of the imperfectly "correlated" type *A*, as Millikan's account proposes, then both sets of *A*-tokens should carry information about *B* (as Shea accepts with respect to his own account). To insist that only the *B*-caused *A*-tokens carry information is to imply that carrying information is not simply a matter of "correlation", after all.[15]

At this point, Scarantino and Piccinini's (2010) distinction between probabilistic and 'all-or-nothing' information becomes relevant. Information that there *is* a visitor at the door is 'all-or-nothing' information, whereas information resulting from (imperfect) probability-raising is probabilistic information, i.e. information that there is a visitor with some increased degree of probability.[16] Now, both wild and normal tokens are said to carry information of this kind (Scarantino and Piccinini, 2010, p. 318). Furthermore, probabilistic information only implies the truth of the probabilistic claim (Piccinini and Scarantino, 2011, p. 23, footnote 25) and that claim is true even when there is no visitor (Scarantino and Piccinini, 2010, p. 316).

Scarantino and Piccinini do not combine these assertions in order to address the issue of wild tokens. But doing so is attractive, because the three claims imply that the wild token ring of a doorbell carries the information that there is a visitor with some increased degree of probability *and* that this information is true even without a visitor. In other words, wild tokens (1) carry natural information and (2) the information is always true; there is no misinformation. The threat of misinformation is circumvented by insisting that probability-raising never delivers 'all-or-nothing' information of the sort that there *is* a visitor; and it is only this type of information that would be false in the case of wild tokens.[17]

4. The significance of probabilistic information

In the preceding section I compared three different ways of articulating natural information in terms of probabilistic relations. This section focuses on probability-raising (as opposed to degrees of coincidence) and asks in the spirit of the first project (see section 1): what is the significance of probability-raising for organisms?

Proponents of probabilistic theories have both a general and a more specific answer to this question. The general answer is that natural information (i.e. probability-raising and, perhaps, simple levels of coincidence) is an important factor in guiding perception, cognition, behaviour, and the evolution of signalling systems (Millikan, 2004, Shea, 2007, Scarantino and Piccinini, 2010, Skyrms, 2010). For instance, natural information is said to play a central role in explaining why

vervet monkeys respond to the alarm calls of their group members in the specific ways they do (Piccinini and Scarantino, 2011, Scarantino, 2013). Proponents also agree that *theories* of natural information should reveal the roles of natural information in particular contexts.

The more specific answer can be gleaned from scattered remarks in the literature. Consider the following statements:

"[Spots] reliably correlate with measles, smoke reliably correlates with fire, and ringing doorbells reliable correlate with people at the door. It is by virtue of these correlations that *one can dependably infer* measles from spots, fire from smoke, and visitors from ringing doorbells." (Scarantino and Piccinini, 2010, p. 318, my emphasis)

"What seems to be happening in these cases [of using natural signs in everyday life] is that there is a real causal connection between two things such that in the circumstances one does depend on the other, and that given what one already knows, *one is able to track* that connection and hence *come to know or suspect* one thing on the basis of knowing the other. […] A natural sign of a thing is something else from which you *can learn* of that thing by tracking in thought a connection that exists in nature. The notion of a natural sign is at root an epistemic notion." (Millikan, 2004, p. 37, my emphasis)

These remarks suggest that probabilistic relations have certain epistemic properties: we can use them in order to draw inferences, come to know and suspect one thing from another. In this sense events can be 'informative' or carry information about other events.

Note that the idea that we "can" make inferences because of probabilistic information is ambiguous. It might mean that probability-raising makes it biologically *possible* for organism to engage in inferences ('can' understood in the modal sense). Or it might mean that probability-raising somehow *entitles* organisms to their inferences ('can' understood in the normative sense of permitting). The terminological ambiguity is unintentional and not much hangs on it. But it is a useful starting point because, as I will argue below, it already encapsulates the two reasons that render probabilistic information significant for the lives of organisms.

Before making the case for this claim it is worth noting that the ambiguity seems to have hidden diverging views about what probabilistic information can achieve, which in turn may have given rise to the implicit disagreement about whether probabilistic relations must be sustained by underlying reasons (Millikan, 2004, Shea, 2007) or not (Scarantino and Piccinini, 2010).

If *A*'s probabilistic relation with *B* is entirely accidental, then there would be no reason to assume that that relation will persist into the future. We may retain the habit of predicting *B* from *A*. But

without an underlying mechanism sustaining the relation we would arguably have no *justification* for predicting *B*. So if one thinks that probabilistic information is important because it provides externalist warrant, then it would be natural to require that the probabilistic relation be non-accidental (but see Lange, 2000). On the other hand, if one values its enabling power, then accidental relations need not be avoided. From the point of view of the inferential abilities of organisms it makes no difference whether a probabilistic relation is accidental or based on underlying causes. Organisms can acquire the ability to engage in inferences from accidental relations because the mechanisms involved, individual learning and evolution, are insensitive to whether or not there is an underlying cause, if so, what it consists in. So long as organisms are exposed to a probabilistic relation and derive a net-benefit from being responsive, they will acquire the disposition to respond. Just how organisms acquire their dispositions, and how probabilistic information figures in these mechanisms, will be the topic of the next section.

4.1 Making inferences biologically possible

Beginning in the late 1960s, learning theorists gradually realised that associative learning normally relies on one event raising the probability of another, a feature known as 'contingency' in learning theory. Experiencing a probability-raising relation can make the difference as to whether animals (including humans) are able to learn from certain stimuli. It is a hard-won insight of learning theory that associative learning depends on the contingency between stimuli, not on their spatial or temporal contiguity. Animals will not learn to predict a shock from a light if they are simply paired together. Only when the light increases the probability of the shock will the animal be able to infer one from the other (e.g. Pearce, 2008, Shettleworth, 2010). Robert Rescorla's early work with rats can serve as an example (as summarized in Rescorla 1988). It also illustrates a point well-familiar to students of animal learning, which is that less-than-perfect probabilistic relations can be sufficient for associative learning (assuming other conditions are satisfied, on which more below).

Rescorla exposed rats to tone-shock pairings in which the tone increased the probability of shock by 40% [P(shock|¬tone) = 0, P(shock|tone) = 0.4]. The rats acquired the maximum strength of association[18] between shock and tone although the tone coincided only occasionally with the shock (40%). An even weaker degree of coincidence (20%) still generated a near-maximum strength of association [P(shock|¬tone) = 0, P(shock|tone) = 0.2]. But when the tone did not increase the probability of shock [P(shock|¬tone) = P(shock|tone) = 0.2], the rats hardly learned to associate the two stimuli.

Findings such as these show that organisms exposed to specific but imperfect probabilistic relations may, through the operation of a learning mechanism and given certain background

conditions, acquire the ability to use that relation to infer[19] one event from the other. Had the rats *not* been exposed to the probability-raising relation between tone and shock, then they would have been unable to acquire that specific inferential disposition. Associative learning is thus an example of probabilistic relations 'allowing' inferences in the modal sense: the tone's raising the probability of shock made it biologically possible for the rats to acquire, and subsequently exercise, a certain inferential ability. 'Biological possibility' is here relative to what members of a certain species can normally do. If there existed a rat that inferred a shock from a tone without prior exposure to the probabilistic relation, simply as a result of some unusual brain process, then it would still be true that normally rats require prior exposure to a probability-raising in order to learn.[20]

A further insight of learning theory is that probability-raising is insufficient for associative learning. This insight puts the significance of natural information into perspective. Consider two additional factors of associative learning, surprise and 'belongingness'. Suppose a rat is first exposed to a noise (CS_1) followed by an electric shock (US) until it has learned to predict the shock from the noise. Subsequently the rat is repeatedly exposed to a noise (CS_1) and a light (CS_2) at the same time, followed by the shock. If after this training the rat is exposed only to the light (CS_2), it does not predict the shock. This is despite the fact that the light increases the probability of the shock. This phenomenon is known as 'blocking': learning to predict the US from one conditioned stimulus, CS_1, blocks the ability to predict the US from the other, CS_2 (e.g. Rescorla, 1988, Pearce, 2008). Experiencing CS_2 during the learning phase yields no suprise: it does not predict anything the rat cannot already predict from CS_1.

'Belongingness' or relevance denotes the fact that many organisms can associate some type of stimuli more easily than others. Most birds can associate illness with colours, but not with flavours. Quails that were offered coloured (blue) water and then lightly poisoned to make them ill learned to avoid blue water, whereas quails that were offered flavoured (and colourless) water did not learn to infer illness from the flavoured water. Learning theorists express this asymmetry between stimuli by saying that, for quails and many other birds, colour 'belongs to', or is relevant for, illness (but flavour is not). By contrast, illness in most mammals 'belongs to' taste, e.g. rats can predict illness from sweetened water, but not from a noise or from flashing lights.[21] Findings like these have undermined the earlier view that animals are capable of connecting any conditioned with any unconditioned stimulus ('equipotentiality of learning', e.g. Rescorla, 1988, Pearce, 2008). The fact that contingency (i.e. probability-raising) is only one of several factors enabling learning is relevant for assessing probabilistic theories as explications of 'information' (see section 4).

We saw that probabilistic relations enable organisms to engage in specific inferences, and it is natural to understand inferences and predictions as involving representational states. On this view,

inferring *B* from *A* amounts to stimulus *A* eliciting a representation of *B* or, alternatively, to stimulus *A* eliciting a representation of *A* that in turn prompts a representation of *B*. The involvement of representational states is in line with cognitive interpretations of associative learning, according to which what is learned is a connection between the conditioned stimulus (e.g. noise) and a neural representation of the unconditioned stimulus (e.g. shock). The resulting behaviour is mediated by the representation ('S-S theories'). But it is likely that not all associative learning relies on representations. Many animals may have a different learning mechanism, i.e. the 'stimulus-response' mechanism of classical conditioning theory. Suppose an animal like a rat has an intrinsic disposition to respond to an unconditioned stimulus (e.g. shock) with an unconditioned response (freezing). The 'classical' mechanism transfers the control of the response to the conditioned stimulus, so that the latter can now elicit the same response as the unconditioned stimulus. What is learned (i.e. acquired) is a connection between the conditioned stimulus and the response the animals gives to the unconditioned stimulus. Associative learning can involve both mechanisms, and in some species both may be present (Shettleworth, 2010). The terms 'predicting' and 'inferring' are therefore best employed in a broad sense so as to include non-representational behaviour that merely *appears as if* the organism had inferred something via a mental representation. The important point for our purposes is that probability-raising plays a role in both cases: it allows organisms to acquire the ability to draw inferences by either mechanism.

We are now in a position to see how exactly probability-raising can be explanatory. I focus on the explanation of behaviour, one of the three phenomena for which natural information is deemed important. Whether a similar approach would work for the other two, cognition and perception, is beyond the scope of this paper. But at least it will be seen how natural information can play a well-delineated and significant explanatory role in one domain of life.

Suppose an organism responds with a type of behaviour *B* to events of type *A*. Since organisms exhibit *B* specifically in response to experiencing *A*, *B* is part of a dispositional property of these organisms. Suppose we wish to explain, in the first instance, why the organism responds to a particular event *a* with a manifestation of behaviour *b*. The explanation for *b* might proceed as follows:

[E 1]

1. Organism *o* has the disposition to respond to *A* with *B*

2. There was an *A*-token (*a*)

Therefore, *o* responded to *a* with behaviour *b*

In other words, the organism's having the disposition, together with perceiving *a*, explains why the organism responded with specific behaviour *b*. But the explanation has limited force; it is like appealing to the melting point of ice when explaining the melting of a particular block of ice. More importantly, it does not involve the *A*-token raising the probability of some other event. It is an ordinary causal explanation; it does not appeal to probabilistic information in any obvious way. So in what sense is the explanation informational?

Let us push our inquiry further by asking why the organism has the disposition in the first place. Assume that the disposition to respond to *A* with *B* is acquired by associative learning. The facts may then be as follows:

[E 2]

1. In the past, the conditioned stimulus *A* increased the probability of the unconditioned stimulus *C*.
2. *C*'s occurrence was surprising.
3. *A*'s physical properties 'belong' to *C*, given the organism's physiology

Therefore, the organisms acquired the disposition to respond to *A* with *B*.

[E2] shows that *A*'s increasing of *C*'s probability is part of the explanation of how the organism acquired the disposition within its lifetime. As we saw above, associative learning is a function of contingency (premise 1), surprise (premise 2), and 'belongingness' (premise 3). There are other factors, but this list will do for my purposes. Since having the disposition is, in turn, part of the explanation of *o*'s token behaviour *b* [E1], the probability-raising between *A* and *C* is part of the full explanation of *b*. Probability-raising (or 'information') explains the organism's token behaviour indirectly, i.e. by explaining why the organism has the disposition to manifest the behaviour in the first place. This is just the sort of result probabilistic theories should deliver: explanations of token behaviour in which natural information is a constituent part.[22]

The preceding discussion of associative learning shows that probability-raising can play a significant role in the lives of organisms. More recent research on probabilistic learning has found that adults and infants are able to quickly respond to statistical regularities in the environment, even if they do not pay specific attention to them and even if the stimuli are not (obviously) biologically salient. This ability, known as 'statistical learning', is thought to be based on computing statistics like transitional probabilities and to enable the delineation of discrete, higher-level units from an otherwise undifferentiated stream of input stimuli (Turk-Browne et al., 2008, Aslin and Newport,

2012). The phenomenon of statistical learning therefore underscores the importance of exploiting probabilistic relations in the environment.

Increasingly, probabilistic (Bayesian) models are used to account for aspects of cognition that cognitive scientists construe as inferences under uncertainty (e.g. Chater et al., 2006, Chater et al., 2011). Classical conditioning, for example, is explained in terms of conditional probabilities involving the US, CS, and the organism's learning history (e.g. Courville et al., 2006). But note that, since Bayesian approaches model *cognitive* processes, the probabilistic relations in question range over degrees of beliefs (Chater et al., 2006), e.g. credences about the events that constitute conditioned and unconditioned stimuli. In this paper, however, I explored the significance of probability-raising relations between these events themselves. After all, the question is whether organisms exploit probabilistic relations that exist in the environment.

So far we considered associative learning as an example of probabilistic relations that 'allow' inferences in the modal sense. Another important mechanism for enabling inferences is evolution. Suppose the members of a population occasionally display a behaviour that can result in exploiting a valuable resource C, but they do so blindly, not knowing whether or not C is present. Suppose also that A positively correlates with C and eventually, thanks to a lucky mutation, some individuals have a heritable tendency to infer C from A. The A-sensitive individuals now have a way to detect C and can therefore exploit C more frequently than the 'blind' ones. On the assumption that this leads to greater lifetime reproductive success, the tendency to predict C from A will spread through the population and eventually become fixed. Here it is the *species* that acquires over many generations the ability to infer one type of event from the other, while particular individuals either have or lack the disposition, depending on their genetic constitution.

Again, the existence of a probabilistic relation between A and C will make a difference as to whether the species will evolve the disposition to infer C from A. Suppose we eliminate probability-raising from this scenario, i.e. $p(C/A) = p(C/\neg A)$. Other things being equal, A-sensitive and 'blind' individuals would then exploit C to the same extent. There would be no benefit for A-sensitive individuals and so their disposition would not spread in the population. This is another example in which the existence of a probability-raising relation makes it biologically possible for organisms to acquire the ability to infer certain things from others (here at species level).

It is easily seen how probability-raising figures in evolutionary explanations of token behaviours. Assume that the disposition to respond to A with B is an evolved response, perhaps a fixed action pattern. The facts may be as follows:

[E 3]

1.   In the past, *A* increased the probability of condition *C*.
2.   Organisms that adjusted their behaviour to *C* in response to perceiving *A*-tokens increased their fitness.
3.   The fitness advantage led to the evolution of the disposition to respond with *B* to *A*.

Therefore, the organism has the disposition to respond with *B* to *A*.

[E3] shows that *A*'s increasing of *C*'s probability (or, equivalently, *A*'s carrying probabilistic information about *C*) in the past partly explains how the organisms acquired the disposition over evolutionary time. Combining this informational explanation of the organism's disposition with [E1] shows that *A*'s carrying probabilistic information about *C* partly explains token behaviour *b* (as did combining [E1] with [E2]).

Note that probabilistic information can explain token behaviour even when the probabilistic relation only holds at type level; it is unnecessary that it should hold at token-level. *Pace* Scarantino and Piccinini (2010), there is no need to regard tokens as the primary bearers of probabilistic information, at least not if the motivation is merely to secure its explanatory role.[23]

In conclusion, there are circumstances in which probability-raising events and objects enable organisms to make predictions. In these circumstances the probabilistic relations are crucial for explaining particular instances of behaviour and for explaining the (species and individual level) acquisition of inferential dispositions. Since probabilistic relations can generate predictions and 'pieces of knowledge' it is natural to describe probability-raising events as 'carrying information' and the probabilistic relations as 'probabilistic information'. There is enough overlap with a colloquial sense of information to motivate informational vocabulary for this class of probabilistic relations. Informational vocabulary picks out probability-raising events in virtue of their consequences for the organisms.

In this section I argued that probability-raising is significant for the lives of organisms because of a *modal* property: it renders certain inferences biologically possible. The next section focuses on a potential *normative* property of probability-raising.

4.2 Warranting inferences

When a state of affairs *A* coincides with another state *B* because of a lawful connection, as Dretske (1981) envisioned, then *B* obtains by nomic necessity (given *A*). So our inference from *A* to *B* cannot fail to be true. Perhaps probability-raising delivers something similar, though less strong. When *A* raises *B*'s probability then *A* makes *B* more likely than it would be otherwise, although it

does not necessitate *B*. In this way *A* may seem to provide some kind of epistemic warrant for inferring *B*. A remark by Millikan points in this direction:

> "From the standpoint of natural epistemology, the cause of one's moving from encounters with *A*s to representations of *B*s […] should be connected with the reason these moves are sometimes correct. Now one's moves of this sort will be based, in central cases, on one's experience […] of a correlation of *A*s with *B*s within some sample. The inference then predicts that the correlation will continue "for a reason" rather than "by accident" only if there is a reason why the correlation persists from the old sample into the new" (Millikan, 2004, pp. 39-40)

As long as there is a mechanism in virtue of which the "correlation" persists, relying on such a "correlation" for one's inferences can yield true beliefs non-accidentally. It is then not implausible to think that reliance on non-accidental probabilistic relations provides a kind of externalist warrant for one's inferences. As a first step in assessing this idea, I focus on three potential sources of externalist warrant: reliabilism, epistemic entitlement, and inductive support.

A natural candidate for warranting inferences based on probabilistic information is reliabilism. Suppose one event (*A*) is a reliable indicator of another (*B*), i.e. if *A* is tokened then *B* is usually tokened as well. Here the relevant notion of 'indication' is the degree of coincidence between *A*s and *B*s. Suppose further that an organism is disposed to infer the condition *B* when experiencing an *A*-token. The organisms' inferences will then usually hold true because the inference '*B* obtains' usually coincides with *B*'s obtaining. In other words, the organism's disposition to draw inferences based on such coincidences is a reliable belief-forming system (or its analogue) in the sense of yielding mostly true beliefs (or their analogues). So it is natural to think that the belief-forming system confers some degree of reliabilist justification on inferences of the form '*B* obtains' and, hence, that the reliable relation itself contributes to *warranting* inferences.

Reliabilist justification is available for Millikan's (2000, 2004) account, according to which *A* carries local information about *B* just in case the degree of non-accidental coincidence between *A*s and *B*s is strong enough to influence sign use. But simple reliabilism of this sort, whatever its intrinsic merits, is unavailable to Shea's (2007) and Scarantino and Piccinini's (2010) theories. Their theories identify information with increases in probability rather than strength of coincidence, and it is easily seen that reliabilism only helps with the latter. Let *A* raise *B*'s probability to a considerable degree, e.g. from 5% to 45%, so that *A* carries a considerable amount of probabilistic information about *B*. Nonetheless, inferring *B* from *A* generates mostly false predictions ($p_{false} = 0.55$). A belief-

forming mechanism that relies on *A*'s correlation with *B* in order to predict *B* is therefore unreliable and does not justify the predictions made. More generally, any probability-raising connections between events that do not yield mostly true inferences lack simple reliabilist justification.

A more promising source of externalist warrant is the idea of epistemic entitlement, particularly Dretske's (2000) version. Dretske argued that one is entitled to hold beliefs that are psychologically irresistible or unavoidable, like the perceptual beliefs we acquire involuntarily when perceiving objects. In a nutshell, if a belief *B* is unavoidable, then it is not within our power to not believe *B*; if we cannot help believing *B*, then we are under no obligation to not believe *B* (assuming that ought implies can); and this implies that we are entitled or permitted to believing *B*. Applied to the present context: once an organisms has acquired the causal disposition to infer or predict *B* from *a*, it cannot help but inferring *B* whenever perceiving *a*, and so it is entitled to predicting *B*. In other words, probability-raising relations can cause organisms to acquire automatic dispositions to infer *B*, and this very fact entitles them to infer *B* from *a*.[24] Thus, Dretske's (2000) account appears to deliver the right result vis-à-vis inferences based on probability-raising events.

Two caveats. First, Dretske (2000) is concerned with the relation between a perception and the corresponding perceptual belief, e.g. the relation between the visual experience as of a tree and the belief that there is a tree. However, probabilistic theories of natural information will usually focus on different relata. They will focus, for instance, on the relation between the perception of a tone (CS) and the non-perceptual belief that a shock will follow (US), not on the relation between the perception of the tone and the corresponding perceptual belief that there is a tone. This is because rats acquire an association between the first, not the second pair of stimuli. So, Dretske's (2000) account of entitlement applies only if *a*'s causing (directly or indirectly) a non-perceptual belief *B* is as unavoidable and involuntary for the organism as is *a*'s causing the perceptual belief *A*. Yet given the 'involuntary' character of much associative learning, this assumption is not implausible.

Second, even granting that probabilistic information confers Dretskean entitlement, one may object that such entitlement does not deliver the required kind of warrant, i.e. something that is conducive to knowledge. The envatted brain, for example, is entitled to its irresistible beliefs in Dretske's sense even though their irresistibility does not render them likely to be true (e.g. Altschul, 2011). Let us therefore turn to an altogether different potential source of externalist warrant, inductive support.

Suppose an organism has encountered *A*-tokens in the past and 80% coincided with environmental condition *C* (perhaps for some underlying reason). Now it encounters a new *A*-token. The organism may then predict, from its past experience and the occurrence of an *A*-token, that there is an 80% probability of *C* occurring. The organism engages in a singular predictive inference:

1. Proportion P = 0.8 of *A*-tokens were accompanied by *C* in the past (for a reason)
2. *a*, so far unobserved, is an *A*

Therefore, the probability of *C* given *a* is 0.8

Another way of putting the conclusion of this argument is to say that an event of type *C probably* occurs, where p(*C*) = 0.8. In other words, *a* and the proportion of past *A*s co-occurring with *C*s provide absolute inductive support for the conclusion (assuming a salient threshold of, say, 0.6). However, absolute inductive support and probabilistic information can diverge. Suppose an event *a* increases *C*'s probability from, say, p = 0.03 to p = 0.3. Since *a* makes an occurrence of *C* much more likely than it otherwise would be (a tenfold increase in *C*'s conditional probability), *a* carries much information about *C*'s occurrence. But *a* does not provide absolute inductive support for the conclusion because *a* does not render *C* likely.

Nevertheless, *a*'s increasing of *C*'s probability means that *a* provides *incremental* inductive support for the belief (or its analogue) that an event of type *C* will occur. Event *a* makes an occurrence of *C* more likely than it would be otherwise, whether or not the total evidence (including *a*) makes *C*'s occurrence probable. The more one event increases the probability of another, the stronger its incremental support for the latter. And conversely, the smaller the rise in probability, the weaker the incremental support. Hence, an event *a* can be said to 'allow' inferring another event *C* in the sense of providing incremental inductive support for that inference. Furthermore, nearly all mechanisms for acquiring an inferential disposition towards certain stimuli involve probability-raising. Apart from one-trial learning, probability-raising is usually required for associative learning, operant conditioning, and evolution. In most circumstances, therefore, organisms cannot acquire a specific inferential ability through these mechanisms unless the latter operate on probability-raising relations. Allowing predictions in the modal sense is therefore generally accompanied by allowing predictions in the normative sense (of incremental support). Bayesian confirmation theory may be used to elaborate the notion of incremental inductive support.[25] Such an elaboration might fit well with Bayesian models of associative learning (e.g. Courville et al., 2006), which are neutral on whether there are explicitly probabilistic computations at the algorithmic and implementational levels (Chater et al., 2006, but see Jones and Love, 2011). To the extent that they do exist at these levels, Bayesian models may not only appropriately describe inferential processes, but also provide a warrant for the inferences they describe. It is unclear, however, how subjective probabilities can be squared with the naturalistic requirements of probabilistic theories of information. These and other

issues about the connection between Bayesian confirmation theory and Bayesian models of inferential reasoning will need to be explored elsewhere.

5 Explicating 'information'

So far I approached probabilistic information in the spirit of the first project outlined above: assessing the role of probability-raising relations for the lives of organisms. We also saw that there is some rationale for referring to probability-raising relations as 'probabilistic information'. But none of this establishes that informational reasoning as practiced in animal behaviour studies or cognitive science consists in reasoning by appeal to probability-raising. Showing that probability-raising has explanatory force for phenomena explored in the behavioural and cognitive sciences neither demonstrates that the information concept, as it is *actually* employed in these domains, amounts to probability-raising, nor that the use of 'information' in these domains is explanatory *in virtue of* referring to probability-raising. Demonstrating this would require a project of the explicatory type, i.e. a detailed analysis of how the information concept is actually employed within a scientific domain, including perhaps the several distinct ways in which it may be used. Proponents of probabilistic information have not provided such explications for any domain. In what follows I describe examples of actual uses of 'information' from two distinct fields, animal learning and animal communication. In neither case does 'information' pick out probability-raising.

In one-trial learning, an animal experiences a one-off coincidence between an *A*-token and a *B*-token and then infers *B* when encountering a second *A*-token. That is, the organism needs no more than a one-off experience in order to acquire a new behavioural response. For instance, a rat that eats a food item *once* and then falls ill (on this occasion) will predict illnes when encountering this kind of food a second time. This phenomenon occurs in humans and many animals and is not restricted to food avoidance.[26] Note that *A* and *B* do not correlate prior to the organism acquiring its ability to predict *B* from *A*; *A* and *B* are merely connected by a one-off coincidence. Since *A* and *B* occur only once, the *A*-token cannot properly be said to raise *B*'s probability (at least not if probability is interpreted in frequentist terms; more on this below).

Nonetheless, it is not unusual to find *A* being described as carrying information about *B*. For instance, hummingbirds are said to learn information about flower locations from a single visit[27] and non-gustatory cues are taken to carry food-related information for two monkey species[28]. Sometimes the term 'information' is not used explicitly. Instead an organism is said to be able to 'associate' the CS with the US on a one-off basis, e.g. mice can associate an odour cue (CS) with milk delivery (US)[29] and humans a food item (CS) with disgust (US)[30]. But on a cognitive interpretation of learning, 'association' is construed informationally.[31] This fact that has been made explicit even in

the case of one-trial learning (see footnote 26). In these examples from the animal learning literature an event *A* is construed, explicitly or implicitly, as 'carrying information' about another event *B* despite *A* not raising *B*'s probability.

It might be objected that token probabilities are in fact admissible on a propensity interpretation of probability, because an *A*-token would have a propensity to bring about a *B*-token. But a propensity interpretation is not applicable in the present context. First, one-trial learning sometimes occurs between causally and nomically unrelated events. Rats can learn to avoid a food item even though the cause of its illness was not the food, but exposure to x-rays (Haley and Snyder, 1964)[32]. On any standard understanding of physical propensity, a token food item in such experiments cannot have a propensity to bring about illness in the rat because the food's properties have no influence whatsoever on whether or not the rat will fall ill. Second, propensity interpretations may account for the conditional probability of a token effect given a token cause (forward conditional probability), e.g. the probability that the ball will hit the net given it has been struck in a particular way. Balls that are struck in certain ways may be thought of as having physical propensities to go in some directions but not others. Yet probabilistic theories rely on *inverse* conditional probabilities, i.e. the probability of a cause given an effect (e.g. the conditional probability that there is a predator given an alarm call). And it is hard to think of the effect as having a physical propensity vis-à-vis its cause (Demir, 2008).[33]

The second type of example comes from the varied use of 'information' in animal communication studies. Depending on the type of signal/organism, researchers use a signal's purported information content interchangeably with (1) what the signal correlates, (2) what a receiver predicts or infers when perceiving the signal, and (3) with the content of the receiver's mental representation that the signal is taken to elicit. These practices suggest that the animal communication literature employs at least three distinct, though interrelated, information concepts. Where 'information' is used interchangeably with the events or things with which a signal correlates, the underlying information concept could be probability-raising. For instance, researchers tacitly assume that the 'dance' of honey bees carries information simply in virtue of being correlated with the location of valuable resources, i.e. independently of whether or not other bees use the correlation when foraging (Stegmann, 2013). Conceivably, the information attributed to the dance is 'food in location X' *because* the dance raises the probability of there being food in location X.[34]

However, where a signal's information content is used interchangeably with what the receivers infer, or with the content of their mental representations, the information concept does not appear to consist in probability-raising. The alarm calls of vervet monkeys can illustrate this point. Vervet monkeys emit three acoustically distinct types of alarm calls in response to three different types of

predators. The information content attributed to calls emitted in response to approaching eagles ('eagle alarm calls') is that an eagle or aerial predator is approaching (Seyfarth and Cheney, 2003). Since eagle alarm calls make an approaching eagle more likely than it would be without the alarm call, Scarantino and Piccinini (2010) conclude that their theory captures the information attributions of ethologists. However, alarm calls raise not only the probability of an eagle approaching, but also of a feathered animal approaching, a two-eyed object approaching, conspecifics running into the bushes, and so on. If the ethologists' information concept did amount to nothing but probability-raising, then one would expect that all these states of affairs are attributed as the signal's content. The fact that only a few correlated states count as the signal's content is readily explained by Seyfarth and Cheney's (2003) practice of using the alarm call's information content interchangeably with the content of the mental representations that the call is taken to elicit in the receiver ('aerial predator approaching'). This practice suggests that a signal S carries information about X just in case S elicits a mental representation of X in the receiver (Stegmann, 2013). The selectivity of content attributions falls naturally out of such a view: the alarm call (S) raises the probability of many facts, including that of an aerial predator approaching (X), but S elicits only a representation of X in the receiver's mind. Hence S is said to carry information about X.[35]

In sum, there is good evidence that 'information' does not always pick out probability-raising in the behavioural sciences. While the examples presented here cannot be easily brushed aside as exceptions or unrepresentative instances, they are compatible with 'information' picking out probability-raising in other areas of behavioural and cognitive science (they are also compatible with probability-raising being a component of one or more important information concepts). However, proponents of probabilistic information would need to make a case for such usages of 'information'. At present, claims to the effect that probabilistic information successfully explicates 'information' in the behavioural and cognitive sciences are premature. The need for analyses of actual usage has been acknowledged in the teleosemantics literature (Neander, 2006).

6 Summary

Probabilistic theories of natural information maintain that one event carries information about another in virtue of a probabilistic relation between them. I compare and assess three variants of this idea, due to Millikan, Shea, and Scarantino & Piccinini. Two points of internal disagreement are worth highlighting. For Millikan, carrying natural information appears to depend on the degree with which one thing coincides with another, whereas both Shea and Scarantino & Piccinini explicitly require that one thing must raise the probability of the other; this is significant because degrees of coincidence do not imply probability-raising. Another difference is that Millikan and Shea, but not

Scarantino & Piccinini, require correlations to be non-accidental. These differences naturally lead to different judgements as to whether any particular event (or state) carries information about another. The differences also emphasize that the probabilistic approach to natural information allows for several concretisations and is still very much in flux.

I distinguish between two philosophical projects about natural information that proponents of probabilistic theories (and others) tend to run together. One type of project explores the significance of probability-raising for the lives of organisms. The other kind of project assesses whether the notion of information actually used in the sciences amounts to probability-raising. The latter can be pursued for its own sake (e.g. when aiming to understand a fundamental concept at work in a science) or in the service of a broader aim (e.g. as a desideratum for theories of natural information).

With respect to the first project I argue for two claims. First, animal learning theorists recognise probability-raising as an important factor in associative learning, and it is therefore clearly significant for organisms. I also show how probability-raising plays a critical role in explaining both token behaviours and the acquisition of inferential dispositions. Second, probability-raising is important for organisms because it provides incremental inductive support for the organisms' inferences.

With respect to the second project I describe two sets of examples in which the information concepts researchers actually employ do not amount to probability-raising. This undermines claims to the effect that probabilistic theories successfully capture the uses of 'information' in the behavioural sciences.

Acknowledgments

References

Alexander, J. J., Audesirk, T. E. & Audesirk, G. J. (1984). One-trial reward learning in the snail *Lymnea stagnalis*. Journal of Neurobiology, 15, 67-72

Altschul, J. (2011). Reliabilism and brains in vats. Acta Analytica, 26, 257-272

Armstrong, C. M., DeVito, L. M. & Cleland, T. A. (2006). One-trial associative odor learning in neonatal mice. Chemical Senses, 31, 343-349

Aslin, R. N. & Newport, E. L. (2012). Statistical learning: from acquiring specific items to forming general rules. Current Directions in Psychological Science, 21, 170-176

Chater, N., Goodman, N., Griffiths, T. L., Kemp, C., Oaksford, M. & Tenenbaum, J. B. (2011). The imaginary fundamentalists: The unshocking truth about Bayesian cognitive science. Behavioral and Brain Sciences, 34, 194-196

Chater, N., Tenenbaum, J. B. & Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. Trends in Cognitive Sciences, 10, 287-291

Cohen, J. & Meskin, A. (2006). An objective counterfactual theory of information. Australasian Journal of Philosophy, 84, 333-352

Courville, A. C., Daw, N. D. & Touretzky, D. S. (2006). Bayesian theories of conditioning in a changing world. Trends in Cognitive Sciences, 10, 294-300

Demir, H. (2008). Counterfactuals vs. conditional probabilities: a critical examination of the counterfactual theory of information. Australasian Journal of Philosophy, 86, 45-60

Dretske, F. (1981). Knowledge and the Flow of Information. (Cambridge, MA: MIT Press)

Dretske, F. (1988). Explaining Behavior: Reasons in a World of Causes. (Cambridge, MA: MIT Press)

Dretske, F. (2000). Entitlement: epistemic rights without epistemic duties? Philosophy and Phenomenological Research, 60, 591-606

Flores-Abreu, I. N., Hurly, T. A. & Healy, S. D. (2012). One-trial spatial learning: wild hummingbirds relocate a reward after a single visit. Animal Cognition, 15, 631-637

Godfrey-Smith, P. (1992). Indication and Adaptation. Synthese, 92, 283-312

Haley, T. J. & Snyder, R. S. (Eds.) (1964). The Response of the Nervous System to Ionizing Radiation. (Boston: Little, Brown & Co)

Jones, M. & Love, B. C. (2011). Bayesian Fundamentalism or Enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. Behavioral and Brain Sciences, 34, 169-231

Kraemer, D. C. (2013). Against "soft" statistical information. Philosophical Psychology [online first, DOI: 10.1080/09515089.2013.785127]

Lange, M. (2000). Natural Laws in Scientific Practice. (Oxford: Oxford University Press)

Laska, M. & Metzker, K. (1998). Food avoidance learning in squirrel monkeys and common marmosets. Learning & Memory, 5, 193-203

Millikan, R. (2000). What has natural information to do with intentional representation? (appendix B). (In R. Millikan (Ed.), On Clear and Confused Ideas (pp. 1-18). Cambridge: Cambridge University Press)

Millikan, R. (2007). An input condition for teleosemantics? A reply to Shea (and Godfrey-Smith). Philosophy and Phenomenological Research, 75, 436-455

Millikan, R. (2013). Natural information, intentional signs and animal communication. (In U. Stegmann (Ed.), Animal Communiction Theory: Information and Influence (pp. 133-146). Cambridge: Cambridge University Press)

Millikan, R. G. (2004). The Varieties of Meaning. (Cambridge, MA: MIT Press)

Neander, K. (2006). Content for cognitive science. (In G. Macdonald & D. Papineau (Eds.), Teleosemantics: New Philosophical Essays (pp. 167-194). Oxford University Press: Oxford)

Pearce, J. M. (2008). Animal Learning and Cognition: An Introduction. (Hove: Psychology Press)

Piccinini, G. & Scarantino, A. (2010). Computation vs. information processing: why their difference matters to cognitive science. Studies in History and Philosophy of Science, 41, 237-246

Piccinini, G. & Scarantino, A. (2011). Information processing, computation, and cognition. Journal of Biological Physics, 37, 1-38

Rescorla, R. A. (1988). Pavlovian conditioning: It's not what you think it is. American Psychologist, 43, 151-160

Rozin, P. (1986). One-trial acquired likes and dislikes in humans: Disgust as a US, food predominance, and negative learning predominance. Learning and Motivation, 17, 180-189

Scarantino, A. (2008). Shell games, information, and counterfactuals. Australasian Journal of Philosophy, 86, 629-634

Scarantino, A. (2013). Animal communication as information-mediated influence. (In U. Stegmann (Ed.), Animal Communication Theory: Information and Influence (pp. 63-81). Cambridge: Cambridge University Press)

Scarantino, A. & Piccinini, G. (2010). Information without truth. Metaphilosophy, 41, 313-330

Seyfarth, R. M. & Cheney, D. L. (2003). Signalers and receivers in animal communication. Annual Review of Psychology, 54, 145-173

Shea, N. (2007). Consumers need information: supplementing teleosemantics with an input condition. Philosophy and Phenomenological Research, 75, 404-435

Shettleworth, S. J. (2010). Cognition, Evolution, and Behaviour. (Oxford: Oxford University Press)

Skyrms, B. (2010). Signals: Evolution, Learning, and Information. (Oxford: Oxford University Press)

Stegmann, U. E. (2013). A primer on information and influence in animal communication. (In U. E. Stegmann (Ed.), Animal Communication Theory: Information and Influence (pp. 1-39). Cambridge: Cambridge University Press)

Suppes, P. (1983). Probability and information (Open Peer Commentary). The Behavioral and Brain Sciences, 6, 81

Turk-Browne, N. B., Scholl, B. J., Chun, M. M. & Johnson, M. K. (2008). Neural evidence of statistical learning: efficient detection of visual regularities without awareness. Journal of Cognitive Neuroscience, 21, 1934-1945

Endnotes

---

[1] Although this sketch of Dretske's position serves for our purposes, it leaves out many important details and ambiguities (see e.g. Millikan, 2000, and Scarantino, unpublished manuscript) and does not reflect the modifications made in subsequent work (e.g. Dretske, 1988).

[2] Probabilistic theories of information have recently been criticized (Kraemer, 2013, Millikan, 2013).

[3] Another recent account of natural information is Cohen and Meskin's (2006) counterfactual theory, which has been discussed before (Demir, 2008, Scarantino, 2008).

[4] Nick Shea's comments on an earlier draft prompted me to draw this distinction.

[5] Assuming a finite frequency interpretation of conditional probability.

[6] This is also how one referee prefers to understand Millikan's condition.

[7] The three-place reading has a further implication. Once organisms employ a correlation between As and Bs they use A to learn something specifically about B; they do not use A to learn something about all the other things with which A correlates. So the local information A carries is specific. It is not about anything A correlates with, nor about what it correlates with most strongly. It is only about those things whose correlation with A was strong enough to have influenced sign use.

[8] "For example, in the case of the honeybee the incoming bee's dance carries correlational information about the location of nectar" (Shea, 2007, p. 426). It is natural to read "the incoming bee's dance" as referring to the token dance of a particular incoming bee (although the phrase can also be interpreted as referring to the type of dance).

[9] Sign use via a history of natural selection plays a different role in Shea's story: it is what turns information into representational content (thereby putting information back into teleosemantics).

[10] We are told in the next sentence: "On this view, a signal s in state F can carry natural information about an object o being G *simply* by raising the probability that o is G..." (Scarantino and Piccinini, 2010, p. 317, my emphasis).

[11] At one point they require that natural information rely on *physical* relations: "Bearers of natural information stand for what they are about by virtue of being physically connected to it. In the absence of the connection, no natural information is carried" (Piccinini and Scarantino, 2010, p. 242). This statement could be interpreted as presupposing some kind of causal, non-accidental connection. However, Scarantino clarified that it should not be interpreted in this way (pers. comm.).

[12] Scarantino confirmed this reading of reliability (pers. comm.). I believe Scarantino & Piccinini's distinction is both significant and applicable in *actual* circumstances, as well. Suppose the footprints increase the probability of quails in forest *F1* by 94%, but in forest *F2* by only 54% (where both *F1*

and *F2* are actual forests). The 94% increase in quail probability in *F1* is reliable in the sense of being strong, but unreliable in the sense of being unstable beyond *F1*. Moreover, the stability sense of reliability matters for organisms. For instance, predators relying on footprints for hunting quail will do worse when moving from *F1* to *F2*.

[13] While writing this paper Scarantino substantially modified his account (unpublished manuscript). My exposition and assessment of his theory concerns the presently published version.

[14] Skyrms (2010, p. 1) ties information to Grice's notion of natural meaning: "Natural meaning [in Grice's sense, U.S.] depends on associations arising from natural processes. I say that all meaning is natural meaning". It is therefore reasonable to take Skyrms' discussion of information in signalling games as concerning natural information.

[15] Another problem is that if wild tokens are physically indistinguishable from normal tokens, users would respond to the first as they would to the second. And this seems to suggest that, from the user's point of view, one is as informative as the other. If wild tokens carried no information from the user's point of view, it would be mysterious why the user would respond to them in the way it does to normal tokens. See also Kraemer (2013).

[16] According to Scarantino and Piccinini (2010, p. 318), a signal carries probabilistic information about an event as soon as it increases its probability somewhat. Probabilistic information does therefore not require raising the event's probability above a threshold that renders the event likely to occur.

[17] The denial of misinformation is compatible with rejecting the veridicality thesis. The latter says (simplified) that if *A* carries information about *B*, then *B* obtains. This thesis is false according to Scarantino and Piccinini (2010): if *A* is a wild token then *A* carries (true) information about *B* (i.e. that *B* occurs with some increased degree of probability) despite *B not* occurring.

[18] The degree of association formed by the animal was measured by the degree to which experiencing the shock suppressed the animals' ordinary activities, in this case bar pressing in order to obtain food. This common measuring procedure is known as suppression conditioning and the resulting measure as the suppression ratio.

[19] I use mentalistic vocabulary here only for convenience. I will return to this point.

[20] A reviewer suggested characterizing the modal sense as making it biologically possible for an organism to engage in *correct* inferences. But this characterization appears to exclude relevant cases. Suppose an organism is trained (by means of a contingency relation) to infer a food item from a certain smell; furthermore, once it has acquired this ability the circumstances are changed such that the smell is no longer predictive of food. Then the organism would for some time, though with

decreasing frequency, engage in false predictions. In this scenario, exposure to a probabilistic relation makes it possible for the organism to engage, for some time, in false inferences.

[21] These asymmetries in the ability to associate stimuli are thought to arise from the organisms' sensory ecology. Internal illness is usually inflicted by food poisoning, and so sensory modalities used to detect food influence the stimuli likely to be associated with illness. Most birds identify food by visual inspection, but most mammals by smelling and tasting it.

[22] Dretske (1988, pp. 96-107) also discusses learning in order to argue that natural information can explain behavior. But there his argument and his conclusions are very different from what is shown here. First, Dretske does not primarily consider the informational relation between two states of the environment, but rather that between an environmental condition and an *internal state* (although elsewhere Dretske often discusses examples of informational relations between two environmental conditions); so Dretske's argument does not establish that *environmental correlations* are explanatory (unless tied in the relevant way to an internal indicator). Second, Dretske refers to 'indication', which for him is a relation that increases to one the probability of the indicated event; whether his argument also works with imperfect probabilistic relations is left open. Third, Dretske's result is contingent on learning by reward (i.e. instrumental learning): the indicator is recruited as a cause of the behavior because otherwise the behavior does not systematically yield a reward. Here I argue that natural information can be explanatory even without behaviors generating a reward (furthermore, unlike instrumental learning, associative learning is disanalogous to evolution by selection in that no fitness-analogous benefit is required). Fourth, the internal indicator's carrying information is explanatory, for Dretske, because it has the *function* to indicate (it was recruited to prompt the behavior because it indicates its success condition). By contrast, the information carried by an environmental condition that is exploited for associative learning is explanatory despite the condition lacking the biological function to be predictive about another event.

[23] There may be other reasons for tying information to tokens, e.g. in order to capture scientific usage of the term 'information'.

[24] Learned dispositions can be lost over time, as a referee rightly observed. But this fact does not undermine Dretskean entitlement as long as the inferential disposition obtains.

[25] In an unpublished manuscript, which I read after writing the present paper, Scarantino develops a detailed account along these lines.

[26] One-trial learning has been documented in species as diverse as snails (Alexander et al., 1984), birds (Flores-Abreu et al., 2012), mice (Armstrong et al., 2006), monkeys (Laska and Metzker, 1998), and humans (Rozin, 1986).

[27] "The behaviour of the birds during experimental stage 2 shows that hummingbirds can encode and retrieve the spatial position within their environment after a single experience. [...] That a rufous hummingbird, while visiting a rewarded flower for the first time, encodes the spatial location and then remembers this *information* to revisit the flower is a remarkable phenomenon when viewed in an ecological context" (Flores-Abreu *et al*., 2012, p. 635, my emphasis). Note that the authors do not identify the precise nature of *A*; presumably the hummingbird uses a complex set of landmarks.

[28] Laska & Metzker (1998, p. 193) mention in their introduction that "nongustatory modalities may convey important food-related *information*" (my emphasis). They then summarise the results of their study by saying that "both squirrel monkeys and common marmosets [are] able to reliably form associations between visual or olfactory cues of a potential food, in the absence of gustatory cues, and its palatability" (p. 198). Thus, the authors take themselves as having shown that non-gustatory cues, which in the introduction they describe as being able to carry food-related *information*, can serve as the basis of one-trial associations with palatability.

[29] "We positively condition neonatal mice to associate arbitrary odorant CSs with a suckling/milk US […] and show that one-trial learning by this method results in conditioned odor preferences for these odorants." (Armstrong *et al.* 2006, p. 344).

[30] Rozin (1986, p. 185) describes his findings on single-trial acquired taste aversions as a "new instance of Pavlovian conditioning in which a disgusting stimulus is the US and a food or other object is the CS." Furthermore, Rozin regarded the results as being compatible with a cognitive interpretation of associative learning: "We have framed the disgust pairing results reported in this paper or by Rozin et al. (1986) in conditioning terms. This is consistent with, or equivalent to, a formulation in terms of the association of ideas, in which the sight of the food gives rise to a disgusting image" (p. 186).

[31] For example, Rescorla (1988) contrasts contiguity and contingency in informational terms: the "modern view of conditioning as the learning of relations sees contiguity as neither necessary nor sufficient. Rather, that view emphasizes the *information* that one stimulus gives about another. [...] conditioning depends not on the contiguity between the CS and the US but rather on the *information* that the CS provides about the US" " (p. 152 and 155, respectively, my emphasis).

[32] My thanks to Mark Sprevak for raising these cases.

[33] A reviewer suggested that single-trial learning can be construed as involving probability-raising if one allows either subjective probabilities or propensities of entire chance set-ups (rather than of individual events). However, subjective probabilities sit ill with the naturalistic aspirations of

theories of probabilistic information, and chance set-ups do not circumvent the problem of inverse conditional probabilities.

[34] This alone is still insufficient to argue that 'information' picks out probability-raising because the dance raises the probability of other facts that are not attributed as the information content of the dance (e.g. it raises the probability that worker bees will depart towards location X). Some explanation is needed as to why these facts are not attributed as information contents. Perhaps ethologists make a distinction between the information a signal *carries* and the information it *conveys* to some receiver (the standard distinction between signal 'message' and 'meaning' points in this direction). Alternatively, 'information' may be reserved for those correlated facts that ethologists regard as explanatory.

[35] One reviewer also suggested that ethologists might well accept that vervet alarm calls carry a diverse set of probabilistic information, if only it were put to them, and that they might conclude that this is how they use 'information' themselves. The latter result is unlikely given that in this case ethologists use 'information' to pick out the content of a mental representation. But even if they responded as suggested, it is far from clear that this procedure would be a suitable methodology for revealing an information concept that ethologists actually employ. Rigorous and extensive discussions of information concepts are rare in the animal communication literature, a fact which has caused disquiet among some practitioners. The lack of rigour could therefore significantly bias the requested self-assessment in favour of any well-articulated and *prima facie* plausible information concept.