Australasian Journal of Philosophy, 2015
http://dx.doi.org/10.1080/00048402.2014.993665

Routledge
Taylor & Francis Group

INFORMATION AS A PROBABILISTIC DIFFERENCE MAKER

Andrea Scarantino

By virtue of what do alarm calls and facial expressions carry natural information? The answer I defend in this paper is that they carry natural information by virtue of changing the probabilities of various states of affairs, relative to background data. The Probabilistic Difference Maker Theory (PDMT) of natural information that I introduce here is inspired by Dretske's [1981] seminal analysis of natural information, but parts ways with it by eschewing the requirements that information transmission must be nomically underwritten, mind-independent, and knowledge-yielding. PDMT includes both a qualitative account of information transmission and a measure of natural information in keeping with the basic principles of Shannon's communication theory and Bayesian confirmation theory. It also includes a new account of the informational content of a signal, understood as the combination of the incremental and overall support that the signal provides for all states of affairs at the source. Finally, I compare and contrast PDMT with other probabilistic and non-probabilistic theories of natural information, most notably Millikan's [2013] recent theory of natural information as non-accidental pattern repetition.

Keywords: information, communication theory, representation, Shannon, Millikan, Bayesianism

1. Introduction

A general theory of natural information, the sort of information that alarm calls carry about predators and facial expressions carry about emotional states, is needed to account for how perception, cognition, and behaviour occur in human and in non-human animals. These capacities are commonly explained in terms of information processing, but the notion of information itself is surprisingly often treated as an unexplained explainer. As a result, claims about the information processing capacities of organisms become metaphors in search of a theory, and the empirical testability of claims involving information is gravely endangered.

In this paper, I introduce a novel theory of natural information, inspired by Dretske's [1981] seminal theory and capable of solving the main problem such theory faces: namely, setting conditions for information transmission that are too stringent. The Probabilistic Difference Maker Theory (PDMT) of natural information that I offer here is defined by three core ideas: (i) natural information must be captured by a three-place relation between signals, states of affairs, and background data, (ii) carrying natural information about a state of affairs simply amounts to changing its probability relative to background data, and (iii) natural information can be quantified using a measure inspired by Shannon's communication theory and Bayesian confirmation theory.

I first summarize Dretske's theory of information and explain why it is not satisfactory as it stands (section 2). Next, I introduce my own probabilistic PDMT account, centred around new definitions of incremental information and informational content (section 3). I then describe the payoffs of PDMT and point to some of its limitations (section 4). Finally, I respond to Millikan's [2013] objections to probabilistic theories of information (section 5) and summarize what has been achieved (section 6).

2. Dretskean Information *Redux*

Dretske's *Knowledge and the Flow of Information* [1981] is an ambitious attempt to formulate a notion of natural information suitable for the reduction of knowledge and other intentional attitudes. Dretske's core definition takes the following form [ibid.: 65; phrasing mine]:

> *Dretske's Natural Information* 1 (DNI_1): A signal r's being G carries the information that s is F = The conditional probability of s's being F, given r's being G (and k), is 1 (but, given k alone, less than 1).

The inclusion of background knowledge k was meant to capture the fact that 'what information is transmitted [by a signal] may depend on what the receiver already knows about the possibilities that exist at the source' [ibid.]. For instance, a signal telling you that John is neither at the gym nor at the office provides you with precise information about John's whereabouts if you already know he is either at the gym or at the bar or at the office, but it does not provide you with the same information if you ignore John's three possible locations.

A standard criticism of DNI_1 is that it cannot provide a reduction of knowledge to information because information is itself defined in terms of background knowledge.[1] The problem I consider most damaging for DNI_1 is that it sets conditions for the transmission of information that are too stringent [e.g. Suppes 1983; Millikan 2004]. Since real world signals do not generally raise the probability of what they carry information about to 1, DNI_1 is inapplicable to the great majority of signals involved in perception, cognition, and behaviour.[2] Call this the *Stringency Problem*.

[1] In response to this criticism, Dretske [1983: 81] argued that background knowledge is 'dischargeable by recursive applications of the definition'. 'Eventually', he continued [ibid.: 87], 'we reach the point where the information received does not depend on any prior knowledge about the source.' To the best of my knowledge, the details of this recursive solution have never been spelled out in any detail.

[2] Dretske [1981] suggested that not everything that can in principle lower the probability of a state of affairs, given the signal, to below 1 should be counted. Many background conditions can simply be assumed to be stable, namely without any relevant alternatives, and will qualify as 'channel conditions'. For instance, with respect to a ringing doorbell, Dretske argued that the integrity of the wires and the absence of a button-pressing demon are channel conditions. Even if we granted that there are channel conditions whose invariance can be assumed, there appear to be common non-channel conditions that make most signals of practical interest equivocal. In the case of the doorbell, the circumstance of naughty kids who press the button and immediately run away would fit the bill, and would lower the probability that a visitor is at the door, given that the doorbell has rung, below 1.

Dretske also formulated a second definition of information [1981: 245; phrasing mine]:

> *Dretske's Natural Information* 2 (DNI_2): A signal r's being G carries the natural information that s is F = There is a nomic regularity such that, given r's being G, s must be F by virtue of a law of nature or logic.

According to DNI_2, r's being G carries the natural information that s is F if and only if it is a law that, if s had not been F, r would not have been G.[3] DNI_2 was introduced as an 'interpretation' of DNI_1 [ibid.]:

> In saying that the conditional probability (given r) of s's being F is 1, I mean to be saying that there is a nomic (lawful) regularity between these event types, a regularity which nomically precludes r's [being G] . . . when s is not F

This quote mischaracterizes DNI_1, which does not require that the conditional probability of s's being F be 1 given r, but instead that it be 1 given r *and k*. Background knowledge drops out of the picture entirely in DNI_2, and the signal r's being G is required to *necessitate* that s is F. Since information is no longer explicitly defined in a k-dependent fashion, DNI_2 is at least in principle a viable candidate for the reduction of knowledge to information.

However, DNI_2 significantly worsens the Stringency Problem faced by DNI_1. DNI_2 demands not only that correlations are perfect between signals and what the signals are about, but also that the correlations are underwritten by laws. If we think of laws as generalizations that are maximally robust under counterfactual presuppositions, that hold with maximal generality across time and space, and that do not mention specific individuals (e.g. Mitchell [2000]; Lange [2000]), precious few instances of transmission of information in the real world would satisfy DNI_2.

Dretske seemed receptive to this line of criticism, eventually replacing the requirement that correlations must be *lawful* with the requirement that they must be *non-accidental*. As he put it, 'there must actually be some condition, *lawful or otherwise*, that explains the persistence of the correlation' [Dretske 1988: 57]. This is an improvement, but it does not go far enough. DNI_1 and DNI_2 still demand that non-accidental correlations be perfect, whereas most real-world correlations grounding the transmission of natural information are imperfect. To solve the Stringency Problem, we must explain how information can be transmitted by virtue of imperfect correlations. This is the task to which I now turn.

3. The Probabilistic Difference Maker Theory of Natural Information

In the most general sense, when we say that a signal carries natural information about a state of affairs what we mean is that the signal *supports* (or

[3] Cohen and Meskin [2006] have carefully developed a version of this account, in which they preserve the counterfactual, but get rid of the requirement that it must be nomically underwritten. For a critical analysis of Cohen and Meskin's [2006] proposal, see Scarantino [2008] and Demir [2008], with a response from Cohen and Meskin [2008].

countersupports) that state of affairs. For example, when we say that an eagle alarm call carries information about eagles, or that John's bared teeth and compressed lips carry information about John's anger, what we mean is that the signals in question support, respectively, the presence of eagles and John's anger. This is what allows recipients to learn about the states of the world from the signals they receive, grounding the intuitive connection between what information a signal carries and what can be learned from it.

My central suggestion is that we apply the resources of *Bayesian confirmation theory* to the understanding of the relations of support and countersupport between *signals* and *states of affairs*, interpreting the former as *evidence* and the latter as *hypotheses*. Bayesians argue that probability can shed light on how evidence bears on hypotheses, and they conceptualize support and countersupport relations as having three places: one for the evidence, one for the hypothesis, and one for background knowledge.[4]

I will use the phrase '*background data*' rather than '*background knowledge*', because I want to emphasize that background data are sets of propositions that may or may not be known by anyone. In some instances of information ascription, background data will be merely hypothetical, whereas in other cases they will be associated with a signal recipient. The explanation of the behaviour of actual recipients demands focusing on background data available to them, but the definition of natural information can remain neutral on the actual or hypothetical nature of background data.

Exploring the nature of information from the vantage point of Bayesian confirmation theory has several advantages. A key one is that this allows us to appreciate that two notions of support have been systematically mixed together in the literature on natural information. Carnap [1950] was the first to notice that when we say that evidence e confirms hypothesis h, relative to background data d, we may mean one of two things:

Evidence e confirms hypothesis h relative to background data d $=$ p(h | e & d) is high.

Evidence e confirms hypothesis h relative to background data d $=$ p(h | e & d) $>$ Pr(h | d).

Carnap called the former *confirmation as firmness* and the latter *confirmation as increase in firmness*. Contemporary Bayesians refer to the former as *absolute confirmation*, and to the latter as *incremental confirmation* [Huber 2005]. To say that a certain hypothesis h is confirmed in the first sense is to say that

[4] The idea that confirmation is a three-place relation between evidence, hypotheses, and background knowledge has been commonplace among Bayesians at least since Good [1967]. Good asked whether a black crow's being observed confirms or disconfirms the hypothesis that all crows are black, and he demonstrated that this depends on background knowledge. Suppose that our background knowledge includes that we inhabit one of two worlds: either world w_1, where there are 100 black crows, no nonblack crows, and 1 million other birds, or world w_2, where there are 1,000 black crows, 1 white crow, and 1 million other birds. The evidence constituted by a black crow's being observed is also evidence that we are in w_2, because we are much more likely to encounter a black crow in w_2 than in w_1. But since in w_2, unlike in w_1, some crows are *not* black, Good [1967] concluded that a black crow's being observed disconfirms the hypothesis that all crows are black, relative to the background knowledge being assumed (it will confirm such an hypothesis, relative to other pieces of background knowledge).

h is highly probable given e and d, whereas to say that it is confirmed in the second sense is to say that adding e to d increases h's probability.

The distinction between these two notions of support affords an instructive interpretation of Dretske's DNI_1. What DNI_1 requires for r's being G to carry the information that s is F (relative to background data d) is that the signal achieves *confirmation as increase in firmness* of the hypothesis that s is F (i.e. p(s is F | r is G & d) > p(s is F | d)) and that it achieves *maximal confirmation as firmness*, because the posterior probability of the hypothesis has to be 1, given the signal and d (i.e. p(s is F | r is G & d) = 1).

To generalize Dretske's analysis with respect to *confirmation as firmness* and *confirmation as increase in firmness*, we must distinguish between the following two notions that DNI_1 blends together:

> *Incremental Natural Information* (INI): r's being G carries incremental natural information about s's being F, relative to background data d, if and only if p(s is F | r is G & d) $\neq$ p(s is F | d).

> *Degree of Overall Support* (DOS): the degree of overall support provided by a signal r's being G about s's being F, relative to background data d, is equal to p(s is F | r is G & d).

The intuition embodied by INI is that signals carry incremental natural information by changing the probability of what they are about, relative to background data. I will say that a signal r's being G carries *positive incremental information* (or provides *incremental support*) about s's being F when p(s is F | r is G & d) > p(s is F | d), and that it carries *negative incremental information* (or provides *incremental countersupport*) about s's being F when p(s is F | r is G & d) < p(s is F | d). The transmission of positive incremental information amounts to the presence of a *positive correlation* between the signal and what the signal is about (relative to background data d), and the transmission of negative incremental information amounts to the presence of a *negative correlation* between the signal and what the signal is about (relative to background data d).[5]

The intuition embodied by DOS is that the overall degree of support provided by a signal about a state of affairs, relative to background data, is measured by how probable such state of affairs is, given the signal and background data—namely, by the posterior probability p(s is F | r is G & d). Dretske focused only on signals that carry positive incremental information and have a degree of overall support of 1. But signals can carry positive incremental information without providing high overall support, because the probability of the state of affairs whose probability they raise could still be low, given the signal and background data.

[5] To say that X and Y are correlated is simply to say that p(X given Y) is higher or lower than p(X). This notion is often referred to as 'contingency' in the psychological literature on learning, and distinguished from the relation of 'contiguity' between X and Y, instantiated when X and Y co-occur in time. Rescorla [1988] demonstrated that two variables can have contiguity without having contingency, and that it is the presence of contingency rather than contiguity that allows for learning about X from Y in a classical conditioning experiment. Stegmann [forthcoming] argues that the notion of correlation is used differently by different authors in the philosophy of information, referring on some usages to 'contingency' (as I do in this paper) and on other usages to 'contiguity' (Stegmann calls the latter 'correlation as degree of coincidence').

Now, whereas p(s is F | r is G & d) is a straightforward measure of the *degree of overall support*, things become more complicated when it comes to measuring the *degree of incremental support* provided by informative signals. The objective of the next section is to develop a viable measure of incremental natural information.

3.1 How to Measure Incremental Natural Information

My starting point is Dretske's [1981] attempt to quantify information, which relied on Shannon's [1948] communication theory, a branch of probability theory developed to solve the problem of reproducing at a receiver a sequence of symbols generated at a source. Consider the following simplified communication system:

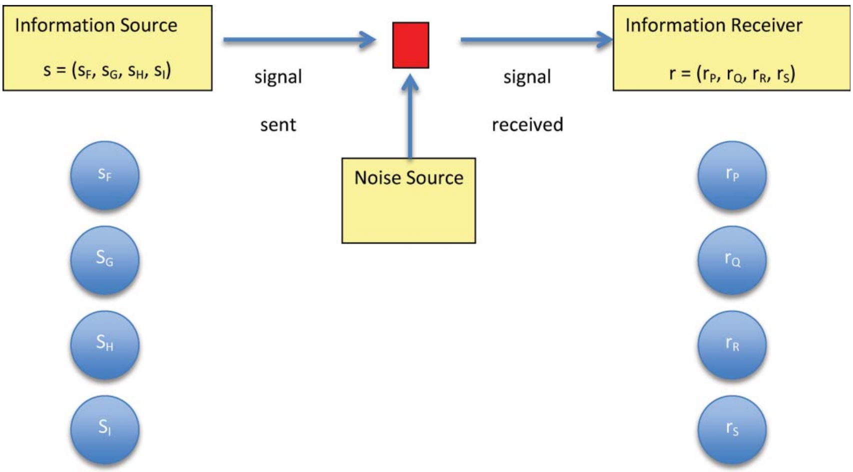

Figure 1. A simplified version of Shannon's [1948] communication system[6]

Let (s_F, s_G, s_H, s_I) be a set of symbols selectable at some source s with probabilities p(s_F), . . ., p(s_I) respectively, and let (r_P, r_Q, r_R, r_S) be a set of symbols selectable at some receiver r with probabilities p(r_P), . . ., p(r_S) respectively. The symbols s_i and r_j represent, respectively, the probabilistic outcomes that 's is i' and that 'r is j', with s and r designating individuals and with i = F, G, H, I and j = P, Q, R, S designating properties. We assume that p(s_i) > 0 for all i = F, G, H, I, and that p(r_j) > 0 for all j = P, Q, R, S, with $\sum_i \mathrm{p}(s_i) = 1$ and $\sum_j \mathrm{p}(r_j) = 1$.

[6] This version is simplified because Shannon's original diagram also represented the processes of encoding symbols into signals to be sent and decoding received signals back into the original symbols.

Shannon [1948] introduced the following two measures:

$$\mathrm{I(s)} = -\sum_i p(s_i) log_2 \mathrm{p}(s_i)$$

$$\mathrm{I(s,\ r)} = \mathrm{I(s)} - \mathrm{I(s \mid r)} = -\sum_i p(s_i) log_2 \mathrm{p}(s_i) - \left[-\sum_j p(r_j) \sum_i p(s_i \mid r_j) log_2 p(s_i \mid r_j) \right]$$

The *entropy* I(s) measures the *average uncertainty* characterizing s, and the *mutual information* I(s, r) measures the *average reduction of uncertainty* characterizing s given r.[7] Shannon obtained the formula for *entropy* by setting a number of mathematical desiderata that any satisfactory measure of uncertainty should satisfy, and by showing that the desiderata could only be satisfied by the formula given above.[8]

The *mutual information* I(s, r) is obtained by subtracting from the entropy I(s), which measures the average uncertainty at the source s *before* the signal is received, the conditional entropy I(s | r), which measures the average uncertainty at the source *after* the signal is received. It is called *mutual* because it is always the case that I(s, r) = I(r, s), namely, the amount of information carried by s about r is equal to the amount of information carried by r about s.

Dretske [1981] realized that neither entropy nor mutual information were suitable measures for his project. This is because they are *average* measures, whereas when we focus on natural information we are interested in the information carried by individual signals. He then proposed the following two non-average measures:

$$\mathrm{I}(s_i) = -\log_2 p(s_i)$$
$$\mathrm{I_S}(r_j) = \mathrm{I}(s_i) - \mathrm{E}(r_j) = \mathrm{I}(s_i) - \left[-\sum_i \mathrm{p}(s_i \mid r_j) log_2 \mathrm{p}(s_i \mid r_j) \right]$$

$\mathrm{I}(s_i)$, labelled as *surprisal*, was defined as a measure of the 'amount of information generated by a particular event or state of affairs [s_i]' [ibid.: 52]. $\mathrm{I_S}(r_j)$ was instead defined as the 'amount of information generated by a particular signal [r_j] about [s_i]' [ibid.]. $\mathrm{E}(r_j)$, finally, is the 'equivocation associated with the particular signal r_j', namely the uncertainty concerning what state of affairs occurred at the source given that r_j was received. I will re-label $\mathrm{I_S}(r_j) = \mathrm{I}(s_i) - \mathrm{E}(r_j)$ as $\mathrm{I_{Si}}(r_j)$ to make explicit the outcomes s_i and r_j with respect to which the amount of communication-theoretic information transmitted is being calculated.

Whereas the measure of surprisal $\mathrm{I}(s_i)$ is in keeping with the basic principles of Shannon's theory, Dretske's $\mathrm{I_{Si}}(r_j)$ fails to measure the mutual information between individual outcomes (see also Lombardi [2005]). What

[7] By choosing base 2 for the logarithm, Shannon committed to measuring information in units equal to the information generated by the occurrence of an outcome with a 0.5 probability ($-\log_2 0.5 = 1$). Since any two possible and equally likely outcomes can be designated with the *binary digits* 1 and 0, the term *bit* was adopted as a name for the unit of communication-theoretic information.

[8] The three mathematical desiderata are the following: (i) The entropy I(.) should be continuous in the probabilities p_i, (ii) the entropy I(.) should be a monotonic increasing function of n when $p_i = 1/n$, and (iii) if $n = b_1 + ... + b_k$ with b_i positive integer, then $I(1/n, \ldots, 1/n) = I(b_1/n, \ldots, b_k/n) + \sum_{i=1}^{k} bi/n\ I(1/bi, \ldots, 1/bi)$.

Dretske overlooked is that the equivocation $E(r_j)$ is still an *average measure*, obtained by taking the sum of $-\log_2 p(s_i \mid r_j)$ for all possible states of affairs s_i at the source, weighted for the conditional probabilities $p(s_i \mid r_j)$. Since $E(r_j)$ is constant, the difference between, say, $p(s_F \mid r_j)$ and $p(s_G \mid r_j)$ does not affect the difference between $Is_F(r_j)$ and $Is_G(r_j)$, preventing Dretske's measure from reflecting the statistical dependencies between individual outcomes.

This problem can be solved by using the following measure [Fano 1961]:

$$I^*(s_i, r_j) = \log_2 \frac{p(s_i \mid r_j)}{p(s_i)}$$

I label $I^*(s_i, r_j)$ as *singular mutual information*. Just as entropy is a weighted average of measures of surprisal, mutual information is a weighted average of measures of singular mutual information. In other words, if we take the sum of all $\log_2 \frac{p(s_i \mid r_j)}{p(s_i)}$, weighted by the probability $p(s_i, r_j)$ of s_i and r_j's joint occurrence, we obtain $\sum_{i=1, j=1}^{n, k} p(s_i, r_j) \log_2 \frac{p(s_i \mid r_j)}{p(s_i)}$, which is Shannon's mutual information $I(s, r)$ under an alternative formulation.

Dretske put $I_{Si}(r_j)$ at the service of a convoluted communication-theoretic argument in favour of his central definition of natural information (DNI_1). In the argument, Shannon's theory plays a modest role: it provides quantitative constraints that need to be satisfied by any definition of the conditions under which a signal r's being G carries the natural information that s is F. Shannon's theory can do much more: namely, it can provide a direct measure of the *amount of incremental natural information* that a signal r's being G carries about s's being F.

Since I have identified incremental natural information with confirmation as increase or decrease in firmness, my measure of incremental information should quantify the degree to which the signal provides incremental support or countersupport for a certain state of affairs. *Singular mutual information* fits the bill. Under the label of *log-ratio measure* (and with background data made explicit), $\log_2 \frac{p(s_i \mid r_j \,\&\, d)}{p(s_i)}$ is one of the measures of incremental confirmation favoured by Bayesians (e.g. Milne [1996]).

As pointed out by Fitelson [1999], any measure c(s is F, r is G, d) of the degree of incremental confirmation provided by r's being G (the evidence e) for s's being F (the hypothesis h), relative to background data d, should satisfy the following *Relevance Requirement*:

$$c(\text{s is F, r is G, d}) \begin{cases} > 0 \text{ if } p(\text{s is F} \mid \text{r is G \& d}) > p(\text{s is F} \mid \text{d}) \\ = 0 \text{ if } p(\text{s is F} \mid \text{r is G \& d}) = p(\text{s is F} \mid \text{d}) \\ < 0 \text{ if } p(\text{s is F} \mid \text{r is G \& d}) < p(\text{s is F} \mid \text{d}) \end{cases}$$

c(s is F, r is G, d) should be positive when r's being G increases the probability that s is F, negative when r's being G decreases the probability that s is F, and equal to zero when r's being G does not affect the probability that s is F (all relative to background data). This requirement is satisfied by $I^*(s_i, r_j)$. I^*(s is F, r is G, d) is positive when p(s is F | r is G & d) > p(s is F | d) because the log of a positive ratio is positive, it is negative when p(s is F | r is G & d) $<$ p(s is F | d) because the log of a negative ratio is negative, and it is equal to zero when p(s is F | r is G & d) = p(s is F | d) because the log of 1 is zero. Provisionally, I conclude that I*(s is F, r is G, d) is a suitable measure of incremental natural information: it is a measure of Shannon's mutual information and is one of the viable measures of Bayesian confirmation.[9]

3.2 Informational Content

I will now focus on the *informational content* of the signal as a whole, understood by Dretske [1981: 47] as 'what-it-is-we-can-learn from that signal' about the possible states of a source. When the signal raises the probability of one state of affairs at the source to 1, the informational content of the signal can be captured by a proposition of the form *that s is F*. Dretske added that 'not all signals … have an informational content that lends itself. . .neatly and economically to propositional expression' [ibid.: 68].

Consider a signal r's being G that changes the conditional probabilities of four equiprobable states at the source s − s is A, s is B, s is C, and s is D, with p(s is A) = p(s is B) = p(s is C) = p(s is D) = 0.25 − as follows:

$$p(\text{s is A} \mid \text{r is G} \,\&\, \text{d}) = 0.07$$
$$p(\text{s is B} \mid \text{r is G} \,\&\, \text{d}) = 0.8$$
$$p(\text{s is C} \mid \text{r is G} \,\&\, \text{d}) = 0.07$$
$$p(\text{s is D} \mid \text{r is G} \,\&\, \text{d}) = 0.06$$

Dretske argued that in this case 'the best we can do is [to say] that the signal carries the information that s is probably in state [B]' [ibid.: 69], because there is a conditional probability of 0.8 that s is B, given the signal and background data. But since we cannot always count on the signal making one specific state of affairs at the source either certain or highly probable, Dretske's proposal should be generalized as follows (assuming the source s can be in one of four states: A, B C, and D):

> *Dretske's Informational Content*: That s is A is overall supported to degree p(s is A | r is G & d), & that s is B is overall supported to degree p(s is B | r is G & d), & that s is C is overall supported to degree p(s is C | r is G & d), & that s is D is overall supported to degree p(s is D | r is G & d).

This account of informational content is predicated on the assumption that the conditional probabilities of the states of affairs at the source are adequate measures of the degree of overall support provided by the signal and background data. Informational contents such as 'that s is B' or 'that s is probably B' correspond to special cases in which the degree of overall

[9] See Floridi [2008] for an overview of mathematical theories of information.

support goes to 1 or close to 1 for s's being B and goes to 0 or close to 0 for all states of affairs other than s's being B.[10]

Now, the problem with this proposal is that we lose track of how the signal *affects* the probabilities of the states at the source. Knowing that the degree of overall support of a state of affairs is, say, p(s is B | r is G & d) does not tell us anything about whether the signal *changed* the probability of s's being B given background data alone. As a result, it may happen that a signal has no effect whatsoever on the probabilities at the source, and yet it has informational content *sensu* Dretske because there exist posterior probabilities, given the signal and background data (see Godfrey-Smith [2012] for further discussion).[11]

A good starting point for a better account is Skyrms's suggestion that 'the informational content of a signal consists in how the signal affects probabilities' [2010: 34]. Skyrms's proposal, adapted to Dretske's example above, goes as follows:[12]

Skyrms's Informational Content:

$$\left\langle log_2 \frac{p(s\ is\ A \mid r\ is\ G\ \&\ d)}{p(s\ is\ A \mid d)}, log_2 \frac{p(s\ is\ B \mid r\ is\ G\ \&\ d)}{p(s\ is\ B \mid d)}, log_2 \frac{p(s\ is\ C \mid r\ is\ G\ \&\ d)}{p(s\ is\ C \mid d)}, log_2 \frac{p(s\ is\ D \mid r\ is\ G\ \&\ d)}{p(s\ is\ D \mid d)} \right\rangle$$

This vector reports the *amounts* of *incremental information* carried by r's being G about the states of affair at the source, quantified by the *singular mutual information* or *log-ratio measure.*[13] Skyrms emphasized that his account, unlike Dretske's own, lacks propositional form. This difference is superficial, however, because we can read propositional content off the vector.[14] A more substantive difference with Drestke's account is that Skyrms takes *informational content* to be defined in terms of *incremental natural information* rather than *overall degree of support*.

[10] In such special cases, the informational content of the signal is as follows: That s is B is overall supported to degree ≈1, and that s is A, s is C, and s is D are each overall supported to degree ≈0. Note that a common way to describe the informational content of signals that do not rule out all states of affairs except one is by means of a disjunction of the form, 'That s is (or might be) either A or B or C or D'. Disjunctions of this form are entailed by the account of informational content I have introduced as a generalization of Dretske's proposal, but they are descriptively much poorer, because they do not specify what the posterior probabilities are for each disjunct, a datum of great relevance to information recipients.

[11] Note that Dretske's first definition of natural information (DNI_1), unlike his second one (DNI_2), includes the requirement that the signal raises the probability of the state of affairs about which it carries information. However, all traces of this requirement are lost in Dretske's propositional rendition of the informational content of a signal, perhaps because Dretske ended up relying mostly on DNI_2 in later parts of the book.

[12] Unlike Skyrms, I have made the background data explicit in my description of the measures of incremental information

[13] Skyrms [2010: 36] also introduced an 'overall measure of information in the signal', understood as the average amount of incremental information carried by the signal about states at the source (this is sometimes called the Kullback-Leibler distance).

[14] As discussed by Birch [forthcoming], Skyrms has a distinctive understanding of the propositional content of a signal, which he understands as the disjunction of the states at the source whose conditional probabilities given the signal are different from zero, namely the set of states that are not ruled out by the signal. On Skyrms's view, then, the propositional content is a special case of informational content, but the latter is descriptively richer because it also specifies the extent to which probabilities have moved. Birch [forthcoming] goes on to argue that Skyrms's notion of propositional content does not allow for signals to have false propositional content.

This is a step in the right direction, because a signal that does not affect the probabilities at the source should not have any informational content with respect to that source. But Skyrms's proposal has two significant flaws. First, we cannot fully capture the informational content of a signal by merely listing the *amounts* of incremental information being carried. Informative signals do not tell us just *how much* probabilities have changed; they also tell us what are the *states of affairs* that had their probabilities changed. On Skyrms's account, two signals that carry the *same amount* of incremental information with respect to four *completely different* states of affairs would have the *same informational content*, captured by, say, the vector $<1, 3.4, -\infty, -1>$. This is unacceptable.[15]

Second, Skyrms does not include posterior probabilities as part of his description of informational content. Godfrey-Smith [2012] has rightly emphasized that recipients are interested in signals to learn what the world is (probably) like, which is what posterior probabilities specify. Godfrey-Smith's view with respect to the two types of informational content anchored by, respectively, posterior probabilities (Dretske's proposal) and changes in posterior probabilities (Skyrms's proposal) is that '[t]here is probably no need to choose one view, saying that such-and-such is the content' [ibid.: 1292].

I share this view, and propose that a full description of the informational content of a signal should report both the changes in probabilities that ground the transmission of incremental information and the final degrees of overall support conferred by the signal, given background data. Accordingly, I define the informational content of a signal with respect to a source that can be in one of n states (state 1, state 2, . . ., state n) as follows:

> *PDMT's Informational Content*: That state 1 is incrementally supported/countersupported to degree $log_2 \frac{p(state\ 1\,|\,signal\ \&\ background\ data)}{p(state\ 1\,|\,background\ data)}$ and overall supported to degree p(state 1 | signal & background data) & . . . & that state n is incrementally supported/countersupported to degree $log_2 \frac{p(state\ n\,|\,signal\ \&\ background\ data)}{p(state\ n\,|\,background\ data)}$ and overall supported to degree p(state n | signal & background data)[16]

This definition includes 'what-it-is-we-can-learn from [a] signal' about a source in its entirety, and is consequently the most complete account of the informational content of a signal. Summing up, what information recipients can learn from signals with respect to the n states of a source includes (i) which states of affairs have had their probabilities changed, (ii) how much the posterior probabilities have changed, and (iii) what are the posterior probabilities. In many cases, the degrees of incremental and overall support will be associated not with precise numbers but with either numeric intervals (degree of incremental support is 0.8ish) or qualitative descriptions (high degree of incremental support). For example, we may describe the informational content of an eagle alarm call as being that the presence of an eagle is incrementally supported to a

[15] This is not a fatal shortcoming, because Skyrms could expand his account of informational content to include, not only the quantities of information being carried, but also an explicit specification of the states of affairs that the information is about. I thank an anonymous referee for emphasizing this point to me.

[16] If no state i is either incrementally supported or incrementally countersupported, the signal should be considered as having no informational content with respect to the source. So the definition I propose makes the implicit assumption that the degree of support/countersupport is different from 0 for at least one of the states of affairs at the source.

high degree and is overall very probable, whereas the presence of another type of predator or of no predator at all, given the call, is countersupported to a high degree and is overall very improbable.

4. What Is the Probabilistic Difference Maker Theory Good For?

4.1 Which Desiderata Does PDMT Satisfy?

Dretske [1983] mentioned several desiderata that a 'philosophically adequate' theory of information should satisfy. First, it should 'preserve enough of our common understanding of information to justify calling it a theory of information' [ibid.: 55]. This common understanding is that 'what information a signal carries is what we can learn from it.' Since I have argued that what information a signal carries is what the signal supports or countersupports, this desideratum is easily satisfied by PDMT, because if a signal X supports or countersupports Y the signal recipient can in principle learn about Y from X. In practice, the signal recipient may fail to learn anything from the signal. What matters is simply that the existence of a support relation between X and Y affords a learning opportunity, which may or may not be exploited by anyone.

Second, Dretske [ibid.] aimed to 'make sense of . . . the theoretically central role information plays in the . . . explanatory efforts of cognitive scientists'. A complication here is that information plays a multiplicity of roles in the explanatory efforts of cognitive scientists. As I have argued elsewhere, the term 'information' designates in the sciences of mind both 'natural information' and what we may call 'non-natural information' [Scarantino and Piccinini 2010]. [17] Natural information is the sort of information that smoke carries about fire, whereas non-natural information is the sort of information that words like 'there is smoke' carry about smoke.

Grice [1957] argued that the key difference between the two types of information/meaning boils down to the possibility of falsehood: whereas there cannot be natural misinformation (smoke cannot carry false natural information about the presence of fire), there can be non-natural misinformation (the words 'there is smoke' can carry false non-natural information about, or misrepresent, the presence of smoke). I am convinced that a naturalistic theory of non-natural information/representation ought to be grounded in natural information, but I will not discuss how to get from the latter to the former in this paper.[18] My objective here is more modest: namely, formulating a theory of natural information that various theories of representation can take as a useful starting point.

What I argue is that PDMT captures the theoretically central role that *natural information* plays in the explanatory efforts of cognitive scientists.[19]

[17] Grice [1957] described the former as natural meaning, and the latter as non-natural meaning.

[18] For a sample of theories of how representations and natural information are related, see, for instance, Dretske [1981, 1986, 1988], Fodor [1990, 1993], Millikan [2000, 2004], Shea [2007], Birch [forthcoming], and Stegmann [forthcoming].

[19] See Stegmann [forthcoming], for an analysis of usages of the information concept that probabilistic theories fail to capture.

Consider the following common uses of the information concept in the sciences of mind:

(1) The eagle alarm call of the vervet monkey carries information about an eagle being present (relative to the background data available to other vervets in the group).

(2) John's frown carries information about his being angry (relative to the background data available to John's wife).

(3) The burning sensation in Mary's fingers carries information about her hand being in contact with something very hot (relative to the background data available to Mary).

(4) The sound of the bell carries information about an electric shock being forthcoming (relative to the background data available to the fear conditioned rat).

(5) The shape of the tracks in the snow carries information about a deer having walked by (relative to the background data available to a deer predator).

(6) The structure of the bee dance carries information about the nectar being 100 feet to the right (relative to the background data available to other bees in the group).

The first point to emphasize is that Dretske's [1981] definitions do not fit any of these cases: the conditional probability of the state of affairs that the signal is about is less than 1, given background data, for all of the examples listed (DNI_1 is not satisfied), and there is no nomic regularity such that, given the signal, the state of affairs that the signal is about must occur by virtue of a law of nature or logic (DNI_2 is not satisfied).

PDMT, on the other hand, fits all of the examples listed, in the sense that signals 1–6 all carry *positive incremental information*: p(state i | signal & background data) > p(state i | background data). In other words, relative to the relevant background data, eagle alarm calls raise the probability of an eagle being nearby, frowns raise the probability of anger, burning sensations raise the probability of contact with a hot surface, sounds of bells raise the probability of electric shocks, the shape of tracks in the snow raises the probability of deer having walked by, and the structure of a bee dance raises the probability that nectar is 100 feet to the right.[20]

[20] This is not to say that, every time scientists utter one of these sentences, what they mean is simply that a change of probability has occurred. In some cases, scientists will mean something more: namely, that the signal has come to carry non-natural information about, or to represent, a certain state of affairs relative to information recipients. Although I will not argue for this point here, my view is that the emergence of a representation relation hinges on the presence of a natural information relation *sensu* PDMT (see Shea [2007] for an argument to this effect). I thank an anonymous referee for pressing me to clarify this point.

4.2 Which Desiderata Does PDMT Not Satisfy?

A third desideratum of Dretske's theory was to reduce knowledge to natural information. Dretske aimed to replace the standard account of knowing that s is F—as truly and justifiably believing that s is F—with an information-theoretic account, according to which knowing that s is F is believing that s is F when such belief is either caused or causally sustained by the natural information that s is F [1981: 85–134].

Finally, Dretske wanted to resist a common tendency among scientists of mind: namely, regarding 'information as a creation of the mind'. He thought that this assumption stood in the way of deepening 'our understanding of the baffling place of mind, the chief consumer of information, in the *natural order* of things' [1983: 55]. A naturalistic notion of information should explain how 'genuine cognitive systems . . . can develop out of lower-order, purely physical information-processing mechanisms' [1981: vii]. This required, for Dretske, accounting for information as a mind-independent commodity that is 'out there, independent of its actual or potential use by some interpreter' [ibid.].

PDMT does not satisfy these last two desiderata. First, given that incremental natural information is transmitted by virtue of changes in conditional probabilities, its reception is generally not sufficient to remove uncertainty completely. Since it is traditionally assumed that knowledge is a 'factive' attitude one can bear only to truths, a belief caused, or causally sustained, by natural information about s's being F does not necessarily constitute knowledge that s is F.[21]

Second, information is transmitted relative to the background data available to hypothetical recipients. On this view, information transmission would not take place in the absence of *potential* interpretative processes on the part of recipients, which violates Dretske's own requirement that information is 'out there' independently of its 'actual or potential use by some interpreter'.

I emphasize that, although mind-dependent, natural information as PDMT understands it should still be understood as an objective commodity, in the same sense in which the confirmation of hypotheses by evidence is an objective affair. Once we have established what the background data are, the natural information transmitted by a signal relative to such data is objectively determined, independently of what anyone may think on the matter.[22]

[21] If we relaxed the conditions for knowledge away from traditional models (e.g. Moss [2013] on probabilistic knowledge), the reception of natural information about s's being F could generate a non-factive variety of knowledge.

[22] Some of Dretske's remarks suggest that this is the notion of objectivity that his theory of natural information was after, rather than the more demanding notion of mind-independence. As he put it [1988: 58–9]:

> [w]e often describe what something means or indicates in a way that reflects what we already know about the possibilities . . . In this sense the meaning we ascribe to signs is relative . . . This, however, doesn't mean that natural meaning is subjective. A person's weight isn't subjective just because it is relative, just because people weigh less on the moon than they do on earth.

Applied to natural information, the idea is that, even if the information carried by a signal is relativized to background data, this does not make it a subjective commodity, provided that the existence of the relevant correlations can be objectively determined. This in turn requires that the probabilities involved in the central definitions of PDMT are given an objectivist interpretation, a point to which I will return shortly.

The fact that PDMT satisfies only two of Dretske's original four desiderata should not be considered a crippling limitation. This is because *no* theory of natural information can satisfy them all. If we conceive of the information carried by a signal in terms of what a suitable recipient can learn from it, as required by the ordinary conception of information, then we cannot conceive of information as a mind-independent commodity. The learning opportunities of recipients change with background data, because what a signal supports or countersupports hinges on background data.

Furthermore, reducing knowledge to information is in tension with building a bridge with the explanatory notion of natural information used in the sciences of mind. This is because the type of information that yields knowledge, as this is traditionally understood, is not the incremental information that scientists of mind commonly invoke to explain perception, cognition, and behaviour of regular organisms in natural environments.

The way forward is to select an internally consistent subset of Dretske's original desiderata and to formulate a theory of natural information that satisfies them. This is the strategy I have pursued in this paper, where I have tried to preserve the intuitive connection between information and learning, and the explanatory significance of information in the sciences of mind, while giving up on reducing knowledge to a mind-independent notion of information. I remain neutral on whether other theories of natural information may also be adequate relative to a different subset of Dretske's original desiderata.

5. Answering Millikan's Objections

Although Millikan [2000] was instrumental to the emergence of probabilistic accounts of information, she has recently given up on them, arguing that they face major problems. She initially distinguished between *nomically underwritten information* (informationL) and *correlational information* (informationC). InformationL is, for Millikan, too rare a commodity for explaining the behaviours of actual creatures in natural environments, so organisms must rely on informationC. Signals [ibid.: 124]

> bearing informationC 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.

This account was recently formalized by Shea [2007: 421], who argued that '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)'.

Give and take a few caveats, this is a special case of INI, in which *correlational information* is identified with *positive incremental information*, a species of the broader genus of incremental natural information. The main caveat is that PDMT does not make any demands on the reasons why the correlation holds, whereas both Millikan [2000] and Shea [2007] require that the correlation holds *for a reason*. I consider this non-accidentality requirement to be unnecessary.[23]

Its inclusion results from the conflation of two projects: clarifying what natural information is, and clarifying when natural information is transmitted reliably.[24] According to PDMT, correlations can be perfectly accidental and yet carry natural information. Take 'all coins in Goodman's pocket are made of silver (at a certain time assumed to be now)', a paradigmatic example of an accidental generalization [Goodman 1947]. Now consider receiving the signal that a certain coin is in Goodman's pocket when your background data include that all coins in Goodman's pocket are made of silver. Under such circumstances, the signal carries the natural information that the coin is made of silver, relative to your background data.

But accidental correlations face two important limitations. First, they do not support counterfactuals: it is false that if a coin had been put in Goodman's pocket it would have been made of silver. Second, they cannot be projected into the future: it is false that, if a coin will be in Goodman's pocket at some future time, it will be made of silver. Accidental correlations are therefore *unreliable*, in the sense that we cannot count on them to persist in future and counterfactual circumstances. Since INI focuses only on providing general conditions for the transmission of natural information, rather than conditions for the *reliable* transmission of natural information, the non-accidentality requirement can be dropped.[25]

I will now consider Millikan's three central objections to correlational information, and I will present some challenges for her new theory of natural information.

[23] Another caveat concerns the way in which the *informational content* of the signal is described. According to Shea's [2007] definition, r's being G ('R') carries the correlational information *that* s is F ('that condition C obtains'), by virtue of the fact that the signal constituted by r's being G makes s's being F more probable than it would be in the absence of the signal. INI holds instead that, under the same circumstances, r's being G carries positive incremental information *about* s's being F. The latter phrasing is preferable. The reason why it is so becomes clear when the degree of overall support provided by s's being F to r's being G is significantly less than 1. Suppose that $p(s \text{ is } F \mid r \text{ is } G) = 0.02$ and that $p(s \text{ is } F \mid \sim (r \text{ is } G)) = 0.01$. According to a plausible interpretation of Shea's definition, r's being G would in this case carry the information *that* s is F, because condition C = s is F had its probability raised by the signal. But saying that the signal carries the information that s is F suggests wrongly that the signal provides *conclusive support* to s's being F. INI's formulation clarifies that what the signal offers is only *incremental support* for s's being F, which is perfectly compatible with a low degree of overall support $p(s \text{ is } F \mid r \text{ is } G)$. Shea could retort that C may also stand for a state of affairs hedged with a probability. On this reading, the information that 'condition C obtains' may be the information that 's is F with low probability.' I prefer my formulation because it is less likely to lead to misunderstandings (see also Kraemer [2015] for a criticism of Shea's [2007] proposal). Finally, Shea [2007] refers to probabilities as *chances*. This is usually a code word for an objectivist interpretation of probability. Shea describes spelling out how to interpret probabilities in an objectivist fashion as a 'topic for another day', but he gives an example—the chance that a lump of uranium-238 will emit an alpha particle in a year—that is often associated with the propensity account. A further possibility is that Shea implicitly endorses Lewis's theory of chances, a sophisticated development of the frequentist interpretation. As I discuss below, PDMT favours objectivist interpretations of probability that are different from both propensity theory and frequentism.

[24] In some of my previous work, I have not been especially clear myself on this distinction (e.g. Scarantino and Piccinini [2010]). See Stegmann [forthcoming], for an analysis of the role of reliability in probabilistic theories of information.

[25] The requirement may well have to be reinstated when we investigate the transition from natural information to representation, because only reliably transmitted information is likely to ground the emergence of representation. I will disregard this point in what follows.

5.1 The Problem of the Single Case

One of Millikan's worries with respect to informationC is that correlations between natural signs, i.e. signs carrying natural information, and the states of affairs that such signs are about, are unnecessary for the transmission of natural information. This is because '[s]ignificant correlations between one thing and another require many repetitions' [2013: 141] and in some cases 'the coincidence of sign with signified occur[s] only once' [ibid.: 137]. Millikan points out, for instance, that the fizzing of a certain bomb carries information about its imminent explosion, and that Suzy's mitten lying on the walk to the side door carries information about Suzy being home from school, even though the informationally related events are not part of any repeated sequence.

These remarks suggest that Millikan takes the probability of an event to be its relative frequency in a repeated sequence of trials. This interpretation is bolstered by her explicit reference to statistical frequencies. She argues that [2004: 32–3]

> [n]early all of the kinds of information needed by us, and by all other organisms as well, for securing what we need in an inclement world, is information that cannot possibly be acquired without leaning on certain merely statistical frequencies.

These statistical frequencies emerge from the repetition over time of events of the same kind.

What Millikan describes as a flaw of the correlational analysis goes under the name of the *problem of the single case*. If probability is relative frequency in repeated trials, the absence of repeated trials deprives events of their probability. Based on this line of reasoning, Millikan concludes that the fizzing of a specific bomb and its explosion are not correlated, and that Suzy's mitten lying in the walk to the side door and Suzy being home from school are not correlated.

Millikan wishes to replace the notion of *correlation* with that of *non-accidental pattern repetition*. Millikan [2013: 140] writes that 'a continuing correlation can be viewed simply as a very small recurring pattern, a repeated pattern of an A state of affairs being so-related to a B state of affairs.' Correlations thus represent one form of pattern. But there are non-correlational patterns too, in which [ibid.: 141]:

> a very detailed complex pattern repeats . . . once in a nearby location', and due to its complexity 'it is likely that the repetition is no accident, the chance of accident going down sharply with the detail and complexity of the pattern.

Millikan's second notion of pattern, however, is also compatible with a correlational account. This is because the existence of a correlation between A and B does not require a repeated sequence of As and Bs. On several *non-frequentist* interpretations of probability, p(B given A) can be different from p(B), i.e. A and B can be correlated, even if there is no sequence of As and Bs.

PDMT favours two objectivist and non-frequentist interpretations of probability: the inductive interpretation and the Bayesian objectivist interpretation. On the *inductive interpretation* (e.g. Carnap [1950]; Maher [2006]), probability is 'ascribed to a hypothesis with respect to a body of evidence' and 'if a certain value holds for a certain hypothesis with respect to a certain evidence, then this value is entirely independent of what any person may happen to think' [Carnap 1950: 43].

The inductive interpretation does not take probabilities to be facts about the world hinging on repetitions of events of the same type, unlike frequentism. Rather, inductive probability is 'relative to the available evidence and does not depend on unknown facts about the world' [Maher 2006]. This is not to say that facts about the world do not affect inductive probability. For example, the relative frequencies of repeatable events can influence inductive probabilities. The point is that they do so *only* if they become part of evidence and/or background data.

Another interpretation of probability that is compatible with PDMT is *objectivist Bayesianism* [Williamson 2009]. Bayesians think of probabilities as degrees of belief, and impose different rationality constraints on them, depending on the specific flavour of Bayesianism that they endorse. Subjective Bayesians demand only that degrees of belief satisfy the axioms of the probability calculus, allowing for major differences in the degrees of belief of equally rational agents faced with the same evidence. This would make it so that the transmission of natural information has to be relativized, not only to evidence and background data, but also to the specific agent that receives the signal, which would turn natural information into a subjective commodity.

Objective Bayesians impose further constraints on rationality, and assume that 'two agents with the same evidence [cannot] disagree as to a probability value and yet neither of them be wrong' [ibid.: 497]. On this interpretation, although probabilities are degrees of belief, once we fix the signal and the background data we have also fixed the degrees of belief that it is rational to have. The important point here is that, if we think of probabilities as rational degrees of belief, there is no obstacle to correlations being objectively defined in the absence of a repeated sequence of events.

I remain neutral in this paper on the choice between inductive interpretations and objective Bayesian interpretations. What is non-negotiable for PDMT's purposes is only that probabilities are relativized to signals and to background data, and that their value is objective once signals and data have been fixed. I conclude that, on both inductive interpretations and objective Bayesian interpretations, the allegedly non-correlational patterns between Suzy's mitten and her being at home, and the bomb's fizzing and its impending explosion, can be captured by a straightforward correlation, dissolving Millikan's first critique.

For instance, the degree of belief it is rational to have that Suzy is at home (or that the bomb is about to explode), given that her mitten is lying by the walk to the side door (or given that the bomb is fizzing), is higher than the degree of belief it is rational to have that Suzy is at home in the absence of her mitten (or that the bomb will explode in the absence of the fizzing). This is all that it takes, on a Bayesian objectivist interpretation, to conclude that Suzy's mitten and her presence at home (or the bomb's fizzing and its impending explosion) are correlated.

5.2 The Reference Class Problem

A second problem that Millikan raises for informationC is the so-called *reference class problem* [Reichenbach 1949]. Just as the *single case problem*, this is a problem historically associated with frequentism. It can be introduced by considering one of Millikan's classic case studies: namely, the behaviour of magnetotactic bacteria, a species of bacteria with the ability to swim along the earth's magnetic field lines by using magnetosomes.

Northern bacteria swim preferentially parallel to the magnetic field, which leads them to seek north and therefore to follow the downward inclination of the magnetic lines in the northern hemisphere. As a result, they swim toward the bottom of the sea, where oxygen is less concentrated. Southern bacteria swim preferentially anti-parallel to the magnetic field, which leads them to seek south and therefore to run counter to the upwards inclination of magnetic lines in the southern hemisphere. This also leads them to swim toward the bottom of the sea, where oxygen is less concentrated.

Consider now the probability that there is less oxygen in direction north, given that the magnetosomes of a given bacterium—let us call him Little John—point north. A frequentist would interpret this conditional probability as a ratio between the number of bacteria in the reference class to which Little John belongs whose magnetosomes point north when oxygen is in direction north and the number of bacteria in the reference class to which Little John belongs whose magnetosomes point north.

The problem is that there are innumerable reference classes to which Little John belongs, which will give rise to different conditional probabilities. In the reference class formed by northern-pointing bacteria in the northern hemisphere, there will be a certain ratio of oxygen-pointing ones, whereas in the reference class formed by bacteria all over the ocean we will get a different one (in the southern hemisphere, less oxygen is in direction south).

Millikan also points out that nothing prevents us from taking a gerrymandered reference class—say, a star of David with a surface area of 2,343 miles, with Little John at the centre—and measuring correlations within such a class. She concludes that '[a]ny single bacterium lies within an infinite number of different designatable areas, but to speak meaningfully of "correlational information" we must decide on some limited reference class' [2007: 450]. Millikan takes this problem to be fatal to Shea's [2007] account of correlational information, which is a special case of my own account (with the caveats I expressed earlier). Since Shea does not specify how to pick a reference class, Millikan [2007: 447] concludes that 'Shea's own description of "correlational information" does not capture anything definite.'

In order to capture something definite, Millikan thinks, a correlational account must be combined with a criterion for picking up a *non-arbitrary* reference class. Her positive proposal is that 'the relevant non-arbitrary reference class consists in the very samples sampled by the animal or species that uses a [signal]' [2013: 138]. I find this suggestion very insightful, and a promising start towards solving the reference class problem. But I note that it marks a key transition from a frequentist interpretation of probability to something very close to the inductive and objective Bayesian interpretations of probability that PDMT favours.

To think of probabilities as being relativized to data about sampled frequencies is a *special case* of thinking of probabilities as being relativized to background data in general. The latter formulation is preferable, because although background data often include data about frequencies in samples, in many cases such data are missing, either because the event is not repeated or because it is not repeatable in principle. Furthermore, data about frequencies, even when present, are not the only data relevant for determining probabilities of states of affairs given signals. Other relevant data may include data about probabilistic propensities, data about conceptual connections (e.g. being square entails having four sides), data about laws of nature (e.g. copper conducts electricity), and innumerable other data about matters of empirical fact (e.g. Suzy's violin recital ends at 3 p.m. today).

The central point is that, even if we restrict the background data to sampled frequencies, Millikan's suggested solution to the reference class problem makes the transmission of natural information to a given recipient contingent upon the background data available to that recipient. As Millikan notes [ibid.], this 'changes the game quite radically … from the one played by theorists of "natural signs", who have hoped to find natural signs and natural information as an 'objective commodity' [Dretske 1981] out in the world'.

It changes it radically because natural signs start resembling 'affordances like food or like shelter, things that are what they are only relative to an animal who would use them' [Millikan 2013: 138]. By the same token, a signal is informative with respect to a given state of affairs only relative to a particular signal recipient and contingently on the background data available to him or her. I have argued in section 3.2 that this does not amount to giving up on the objectivity of natural information, but to understanding such objectivity as being relativized and mind-dependent. As we think of a tree as being objectively climbable, relative to a given squirrel, or of the velocity of a car as being objectively 60 m.p.h., relative to a certain observer, so we should think of a signal as being objectively informative, relative to a given signal recipient.

In conclusion, PDMT solves the reference class problem by acknowledging openly that one cannot determine whether signals and states of affairs correlate *simpliciter*, but only whether they correlate *relative to a specific set of background data*. To figure out the behaviour of actual signal recipients, we should follow Millikan's astute suggestion and focus on the background data—about sampled frequencies and other relevant data—that are available to the organisms that use the signal.

5.3 The Veridicality Problem and A Challenge for Millikan's Account

Millikan's final worry is that informationC can be carried about states of affairs that do not occur (see also Floridi [2007]). This is indeed an implication of PDMT, as well as of Shea's [2007] and Skyrms's [2010] probabilistic theories of information. If r's being G carries incremental information about s's being F just in case p(s is F | r is G & d) ≠ p(s is F | d), then whether or not s is F is irrelevant to the transmission of natural information about s's being F. As a result, it may be the case that r's being G carries positive incremental information about s's being F, despite the fact that s is not F, and that r's being G carries negative incremental information about s's being F, despite the fact that s is F.

Millikan finds this consequence objectionable (see also Kraemer [2015] and Stegmann [forthcoming]). On her view, the fact that s is not F should prevent a signal from carrying natural information about s's being F (and presumably the fact that s is F should prevent a signal from carrying natural information about s's not being F). Millikan hints at this requirement in the way her information claims are phrased. She writes, for instance, that [2013: 134; emphasis mine] '[b]lack clouds are *often* a sign of immanent rain' or that 'fever is *sometimes* a sign of measles, other times of flu, and so forth.'

The assumption here is that, in order for black clouds to carry natural information about rain, a pattern between black clouds and rain is not sufficient: rain must also occur. Similarly, what determines whether fever is a natural sign of measles or flu is not merely the existence of a pattern between, respectively, fever and measles or fever and flu, but also whether the patient actually has measles or flu. This adds a veridicality requirement to Millikan's [ibid.] theory of natural information, which I understand as having the following form:

> *Millikan's Natural Information* (MNI): A signal r's being G carries natural information about, or is a natural sign of, s being F = There is a non-accidental pattern relating r's being G and s's being F, & s is F.

I have discussed the notion of a *non-accidental pattern*, suggesting that *pattern* is just another name for *correlation* on the non-frequentist interpretations of probability that I favour, and that the *non-accidentality* requirement is superfluous when we are trying simply to establish conditions under which natural information is transmitted, reliably or not. What I want to focus on here is the further assumption that information about s's being F can be carried only if s is F.

I acknowledge that there is an important sense in which information entails truth. It is well captured by this passage: 'What information a signal carries is what it is capable of "telling" us, telling us *truly*, about another state of affairs' [Dretske 1981: 44]. If so, a signal cannot carry the information that a certain state of affairs occurs if it does not occur. But it does not follow that a signal cannot carry *any* natural information about non-occurrent states of affairs. First, a signal can tell us *truly* that the state of affairs in question is more or less probable than it was before the signal. For example, even if no eagle is present, an eagle alarm call can tell us truly that the presence of an eagle is more probable than it was given background data alone. Second, a signal can tell us *truly* that the state of affairs in question is probable to a degree determined by its posterior probability, given the signal. For example, even if no eagle is present, an eagle alarm call can tell us truly that the presence of an eagle is very probable, given the signal and background data.

Millikan's commitment to the assumption that it is necessary for s to be F in order for a signal to carry natural information about, or be a natural sign of, s's being F leads to unpalatable consequences. Suppose that a person has a probability of 0.1 of having dengue, a probability of 0.1 of having measles, and a probability of 0.1 of having the flu. Suppose also that the presence of fever is strongly negatively correlated with the presence of dengue, strongly positively correlated with the presence of measles, and minimally positively correlated with the presence of flu. For instance, assume that the following values hold: p(dengue | fever) = 0.0001, p(measles | fever) = 0.99, and p(flu | fever) = 0.1001.

Since the three correlations described instantiate three kinds of non-accidental patterns, what determines whether fever is a natural sign of dengue, flu, or measles depends on whether the patient actually has dengue, flu, or measles. This means that the nature of the correlation—as long as there is one—becomes *irrelevant* to establishing *which* natural information is carried by the signal. This consequence is made explicit in the following passage [Millikan 2013: 136]: '[N]o matter how high the non-accidental correlation between fever and measles is in the area, if Johnny's fever is caused by flu then his particular fever would not be, for our purposes, a correlational sign of measles.'

Suppose now that the patient has dengue. By MNI's lights, we would have to conclude that fever carries information about dengue, even though the presence of fever makes dengue *less probable*. After all, there is a non-accidental pattern relating having fever and having dengue, and the patient has dengue. This is clearly unacceptable, because a natural sign of a state of affairs should not make such state of affairs less probable than it would have been had the sign not occurred.

This problem can be fixed. What we need is to require that, in order for a signal r's being G to carry natural information about s being F, it must be the case that r's being G and s's being F are *positively correlated*, not just that they instantiate a non-accidental pattern. This suffices to exclude the possibility that fever is a sign of dengue when it makes dengue less probable. But it would leave all positive correlations on a par, whatever their strength may be. This is also problematic, because the extent to which X is a natural sign of Y should depend on the degree to which X incrementally supports Y.

Suppose now that the patient has the flu. In such a case, Millikan's MNI would have us conclude that fever carries information about the patient having the flu, even though fever increases only imperceptibly the probability that the patient has the flu (from 0.1 to 0.1001), whereas it boosts dramatically the probability that he or she has measles (from 0.1 to 0.99). This separation of the natural information carried by a signal from what the signal incrementally supports has negative consequences for the theory's ability to satisfy the two desiderata that PDMT takes to be non-negotiable: the link with learning, and the preservation of the central explanatory role that information plays in the sciences of mind.

First, Millikan's [2013] theory does not preserve the core intuition that 'what information a signal carries is what we can learn from it.' What we can learn from the patient's fever is that it is extremely probable that the patient has measles, and that their probability of having the flu is barely affected at all. But what natural information the signal carries is that the patient has the flu, which is something that no signal recipient could possibly learn from the signal. On the other hand, fever would not be a natural sign of measles by MNI's lights, despite the fact that signal recipients can learn about the presence of measles with 99% certainty once they observe the fever.

Second, Millikan's theory is in danger of being at odds with the way in which scientists of mind conceive of information in their explanatory practices. I will focus briefly on the role played by information in classical conditioning (but there are lots of other examples). Suppose that in a conditioning experiment the ringing of the bell provides a major boost to the probability of getting an electric shock, making the latter almost certain, but that in extremely rare cases, say once every thousand trials, the ringing of the bell is followed by the release of sugar pellets rather than electric shocks. Suppose now that the bell rings for the conditioned rat, and that sugar pellets are released.

On Millikan's theory, the ringing bell would in such case be a natural sign of sugar pellets. But this is not reflected by the behaviour of the rat, who immediately freezes upon hearing the bell. The reason why this occurs is that the bell provides massive incremental support for the hypothesis that an electric shock is forthcoming. According to PDMT and to the majority of learning theorists in psychology, this is all that is required for the ringing bell to carry natural information about an electric shock. If 'the shock is equally likely whether or not the tone is present', on the other hand, then 'the tone provides no information' [Rescorla 1988: 152].

The problem for Millikan's [2013] theory is that the rat's behaviours are not affected by what signals are supposed to be a natural sign of (sugar pellets), whereas they are affected by what they are not supposed to be a natural sign of (electric shocks). This counterintuitive consequence results from making the actual occurrence of a state of affairs the primary determinant of whether a certain signal carries natural information about it, independently of the strength of the positive correlation between signals of that type and states of affairs of that type. This severs the connection between natural information and incremental support, which is what a theory of natural information must preserve in order to capture the central explanatory role of natural information in the sciences of mind.

I conclude that, contrary to Millikan's [2013] proposal, the ringing bell in the conditioning experiment is *always* a sign of electric shock, black clouds are *always* a sign of imminent rain, and fever is *always* a sign of measles (on the probabilistic assumptions I described earlier), whether or not, respectively, electric shocks, rain, or measles actually occur. They are signs by virtue of how they change probabilities, which in turn affects what recipients can learn from signals and how recipients behave upon receiving the signals.

6. Conclusion

The Probabilistic Difference Maker Theory (PDMT) of natural information is defined by three core ideas: information is a three-place relation between signals, states of affairs, and background data; carrying information about a state of affairs is changing its probability; and natural information is quantifiable, using a measure inspired by Shannon's communication theory and Bayesian confirmation theory. PDMT also includes a novel account of the informational content of signals, developed by combining insights from Dretske [1981] and Skyrms [2010]. PDMT captures the intuitive connection between information and learning, it sheds light on the explanatory role information plays in the sciences of mind, and it circumvents the challenging objections to probabilistic theories of information that have recently been raised by Millikan [2013].[26]

Georgia State University, USA

References

Birch, J. forthcoming. Propositional Content in Signalling Systems, *Philosophical Studies*.
Carnap, R. 1950. *Logical Foundations of Probability*, Chicago: University of Chicago Press.
Cohen, J. and A. Meskin 2006. An Objective Counterfactual Theory of Information, *Australasian Journal of Philosophy* 84/3: 333–52.
Cohen, J. and A. Meskin 2008. Counterfactuals, Probabilities and Information: Response to Critics, *Australasian Journal of Philosophy* 86/4: 635–42.
Demir, H. 2008. Counterfactuals vs. Conditional Probabilities: A Critical Analysis of the Counterfactual Theory of Information, *Australasian Journal of Philosophy* 86/1: 45–60.
Dretske, F.I. 1981. *Knowledge and the Flow of Information*, Cambridge MA: The MIT Press.
Dretske, F.I. 1983. Précis of *Knowledge and the Flow of Information*, *Behavioral & Brain Sciences* 6: 55–90.
Dretske, F.I. 1986. Misrepresentation, in *Belief: Form, Content, and Function*, ed. R.J. Bogdan, Oxford: Clarendon Press.
Dretske, F.I. 1988. *Explaining Behavior: Reasons in a World of Causes*, Cambridge MA: The MIT Press.
Fano, R.M. 1961. *Transmission of Information: A Statistical Theory of Communications*, Cambridge, MA: The MIT Press.
Fitelson, B. 1999. The Plurality of Bayesian Measures of Confirmation and the Problem of Measure Sensitivity, *Philosophy of Science* 66/3: S362–78.
Floridi, L. 2007. In Defence of the Veridical Nature of Semantic Information, *European Journal of Analytic Philosophy 3*/1: 31–42.
Floridi, L. 2008. Semantic Conceptions of Information, *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL = http://plato.stanford.edu/archives/win2008/entries/information–semantic/>.
Fodor, J.A. 1990. *A Theory of Content and Other Essays*, Cambridge MA: The MIT Press.
Fodor, J.A. 1993. *The Elm and the Expert: Mentalese and Its Semantics*, Cambridge MA: The MIT Press.
Godfrey-Smith, P. 2012. Review of Signals: Evolution, Learning and Information, *Mind* 120/480: 1288–97.
Good, I.J. 1967. The White Shoe is a Red Herring, *The British Journal for the Philosophy of Science* 17/4: 322.

[26] This paper has been many years in the making, so I am afraid that I have forgotten some of the people who gave me useful feedback on it. But I have not forgotten them all. The following people deserve my gratitude for either feedback on previous drafts or helpful discussions on some of the topics covered in the paper: Fred Adams, Ingo Brigandt, Jonathan Cohen, Hilmi Demir, Alan Hájek, Douglas Kutach, Aaron Meskin, Ruth Millikan, Eddy Nahmias, Karen Neander, Michael Owren, Gualtiero Piccinini, Ulrich Stegmann, and two anonymous referees for this Journal. I want to emphasize in particular my intellectual debt to Ruth Millikan, who has been a great source of inspiration for my work on natural information over many years.

Goodman, N. 1947. The Problem of Counterfactual Conditionals, *The Journal of Philosophy* 44/5: 113–28.
Grice, H.P. 1957. Meaning, *The Philosophical Review* 66/3: 377–88.
Huber, F. 2005. What is the Point of Confirmation?, *Philosophy of Science* 72/5: 1146–59.
Kraemer, D.M. 2015. Against 'Soft' Statistical Information, *Philosophical Psychology* 28/1: 139–47.
Lange, M. 2000. *Natural Laws in Scientific Practice*, New York: Oxford University Press.
Lombardi, O. 2005. Dretske, Shannon's Theory and the Interpretation of Information, *Synthese* 144/1: 23–39.
Maher, P. 2006. The Concept of Inductive Probability, *Erkenntnis* 65/2: 185–206.
Millikan, R.G. 2000. *On Clear and Confused Ideas: an Essay About Substance Concepts*, Cambridge, UK: Cambridge University Press.
Millikan, R.G. 2004. *Varieties of Meaning: The 2002 Jean Nicod Lectures*, Cambridge MA: The MIT Press.
Millikan, R.G. 2007. An Input Condition for Teleosemantics? Reply to Shea (and Godfrey-Smith), *Philosophy and Phenomenological Research* 75/2: 436–55.
Millikan, R.G. 2013. Natural Information, Intentional Signs and Animal Communication, in *Animal Communication Theory: Information and Influence*, ed. U. Stegmann, Cambridge: Cambridge University Press: 133–48.
Milne, P. 1996. log(P(h/eb)/P(h/b)) Is the One True Measure of Confirmation, *Philosophy of Science* 63/1: 21–6.
Mitchell, S.D. 2000. Dimensions of Scientific Law, *Philosophy of Science* 67/2: 242–65.
Moss, S. 2013. Epistemology Formalized, *The Philosophical Review* 122/1: 1–43.
Reichenbach, H. 1949. *The Theory of Probability*, Berkeley: University of California Press.
Rescorla, R.A. 1988. Pavlovian Conditioning. It's Not What You Think It Is, *American Psychologist* 43/3: 151–60.
Scarantino, A. 2008. Shell Games, Information, and Counterfactuals, *Australasian Journal of Philosophy* 86/4: 629–34.
Scarantino, A. and G. Piccinini 2010. Information Without Truth, *Metaphilosophy* 41/3: 313–30.
Shannon, C.E. 1948. A Mathematical Theory of Communication, *The Bell System Technical Journal* 27: 379–423, 623–56.
Shea, N. 2007. Consumers Need Information: Supplementing Teleosemantics With an Input Condition, *Philosophy and Phenomenological Research* 75/2: 404–35.
Skyrms, B. 2010. *Signals: Evolution, Learning, & Information*, Oxford: Oxford University Press.
Stegmann, U. forthcoming. Prospects for Probabilistic Theories of Information, *Erkenntnis*.
Suppes, P. 1983. Probability and Information, *Behavioral and Brain Sciences* 6: 81–2.
Williamson, J. 2009. Philosophies of Probability, in *Philosophy of Mathematics*, ed. A.D. Irvine, Amsterdam: North Holland: 493–533.