

Chapter 2

The Fundamental Properties of Information–Carrying Relations

Hilmi Demir
Bilkent University, Turkey

ABSTRACT

Philosophers have used information theoretic concepts and theorems for philosophical purposes since the publication of Shannon’s seminal work, “The Mathematical Theory of Communication”. The efforts of different philosophers led to the formation of Philosophy of Information as a subfield of philosophy in the late 1990s (Floridi, in press). Although a significant part of those efforts was devoted to the mathematical formalism of information and communication theory, a thorough analysis of the fundamental mathematical properties of information-carrying relations has not yet been done. The point here is that a thorough analysis of the fundamental properties of information-carrying relations will shed light on some important controversies. The overall aim of this chapter is to begin this process of elucidation. It therefore includes a detailed examination of three semantic theories of information: Dretske’s entropy-based framework, Harms’ theory of mutual information and Cohen and Meskin’s counterfactual theory. These three theories are selected because they represent all lines of reasoning available in the literature in regard to the relevance of Shannon’s mathematical theory of information for philosophical purposes. Thus, the immediate goal is to cover the entire landscape of the literature with respect to this criterion. Moreover, this chapter offers a novel analysis of the transitivity of information-carrying relations.

INTRODUCTION

Philosophers have used information theoretic concepts and theorems for philosophical purposes since the publication of Shannon’s seminal work,

“The Mathematical Theory of Communication”. The efforts of different philosophers led to the formation of Philosophy of Information as a subfield of philosophy in the late 1990s (Floridi, in press). Although a significant part of those efforts was devoted to the mathematical formalism of information and communication theory, a thor-

DOI: 10.4018/978-1-61692-014-2.ch002

ough analysis of the fundamental mathematical properties of information-carrying relations has not yet been done. This is an important gap in the literature because fundamental properties such as reflexivity, symmetry and transitivity are not only important for mathematical purposes, but also for philosophical purposes. For example, in almost all attempts to use information theoretic concepts for philosophical purposes, information-carrying relations are assumed to be transitive. This assumption fits our intuitive understanding of information. On the other hand, the transitivity assumption has some controversial consequences. For information theoretic concepts to be useful for philosophical purposes, the semantic informational content of a signal needs to be uniquely identified. In standard accounts, the informational content of a signal is defined by conditional probabilities. However, conditional probabilities obey transitivity only if when they are 1, and thus the informational content of a signal is fixed in an absolute manner. This leads to the denial of partial information and misinformation, which sounds implausible at first glance (Lehrer & Cohen 1983; Usher 2001). Some have preferred to accept the dichotomy and live with the ensuing seemingly implausible consequence (Dretske 1981). Others have tried to avoid the implausible consequence by using some other notions from the stock of mathematical theory of communication, such as mutual information (Usher 2001; Harms 1998). The point here is that a thorough analysis of the fundamental properties of information-carrying relations will shed light on some important controversies. The overall aim of this chapter is to begin this process of elucidation. It therefore includes a detailed examination of three semantic theories of information: Dretske's entropy-based framework, Harms' theory of mutual information and Cohen and Meskin's counterfactual theory. These three theories are selected because they represent all lines of reasoning available in the literature in regard to the relevance of Shannon's mathematical theory of information¹ for

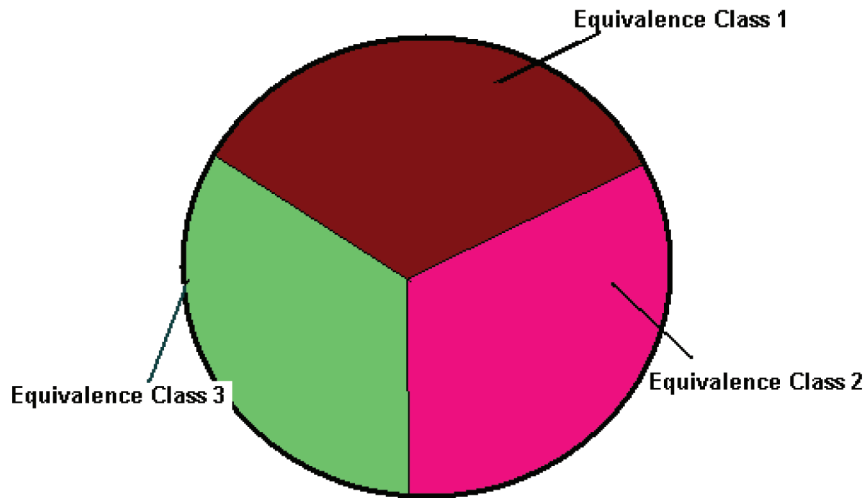
philosophical purposes. Thus, the immediate goal is to cover the entire landscape of the literature with respect to this criterion. Moreover, this chapter offers a novel analysis of the transitivity of information-carrying relations. Until recently, transitivity has been assumed without question. Cohen and Meskin's work (2006) is the first in the literature that challenges this assumption. They claim that information-carrying relations need not be transitive; there are cases where this assumption fails. They state this claim, however, without giving any argument; they simply assert it, which is understandable given the scope of their article. This chapter provides a novel argument in support of their claim. The argument is based on the Data Processing Inequality theorem of the mathematical theory of information.

Given this framework, the chapter is organized as follows. Section 1 is a basic introduction to equivalence relations and may be bypassed by those who are already familiar with this topic. Section 2 is a brief historical survey of the literature. Section 3 analyzes the three semantic theories mentioned in the previous paragraph, in chronological order. Section 4 answers the following question: What are the desired properties of information-carrying relations for philosophical purposes? Lastly, Section 5 concludes the chapter with some suggestions for future research. There is also a short glossary of technical terms at the end.

EQUIVALENCE RELATIONS: A PRELIMINARY INTRODUCTION

A relation could have any number of arguments: one, two, three, four and so on. For example, a 'being in between' relation requires three arguments, that a is in between b and c , and therefore is a 3-place relation. Similarly, 'being the father of' is an example of a 2-place relation with two arguments: the father and the child. These 2-place relations are also called binary relations. Our main focus in this chapter is binary relations, since an

Figure 1. Domain Set



equivalence relation is a binary relation with some specific properties.

A binary relation is a collection of ordered pairs. The first member of the pair comes from the domain; the second member is a member of a set called the co-domain. The collection of ordered pairs is a subset of the Cartesian product of the domain and the co-domain. For example, $A \times A$ is the Cartesian product of A with itself, and any subset of this product is a binary relation. In formal notation, let X (domain) and Y (co-domain) be sets; then any R that satisfies the following condition is a binary relation.

$$R \subseteq X \times Y \text{ or equivalently, } R = \{ \langle a, b \rangle \mid a \in X \text{ and } b \in Y \}$$

When a pair forms a relation, the first member of the pair is related to the second member of the pair. In other words, if $\langle a, b \rangle \in R$, then we write aRb . One of the first questions that mathematicians ask about a binary relation is whether or not it is an equivalence relation. Equivalence relations split up their domain into disjoint (mutually exclusive) subsets; they partition their domain. Each of these disjoint subsets is called an equivalence class. Members of an equivalence class are equivalent

under the terms of the relation. Any member of the domain is a member of one and only one equivalence class. This is a desirable feature because it neatly organizes the domain and avoids any ambiguity in terms of class membership. Figure 1 is a visual example of how an equivalence relation may divide up the domain set into mutually exclusive subsets.

For a binary relation to be an equivalence relation, it has to have three properties: reflexivity, symmetry and transitivity. Reflexivity simply means that every member of the relation has to enter into relation with itself. Symmetry, which may also be called mirroring, implies that for every pair $\langle a, b \rangle$ that falls under the relation, the mirror image of the pair, $\langle b, a \rangle$, is also a member of the relation. As the name suggests, transitivity requires that if a is related to b and b is related to c , then a must be related to c . It may be helpful to state these properties in mathematical notation and explain some of their features.

Reflexivity

A relation is reflexive if and only if

$$\forall a \in X, aRa$$

The Fundamental Properties of Information-Carrying Relations

An example of a reflexive relation is ‘being divisible by itself’. The opposite of reflexivity is anti-reflexivity. Anti-reflexivity is not simply a failure of reflexivity; rather, it is a stronger condition. A relation is anti-reflexive if and only if no member of the domain enters into relation with itself.

A relation is anti-reflexive if and only if

$$\forall a \in X, \sim aRa$$

‘Being the father of’ is such a relation, because no human being is the father of himself. Since anti-reflexivity is stronger than the mere failure of reflexivity, we have some relations that are neither reflexive nor anti-reflexive. For such relations, both the reflexivity and anti-reflexivity conditions fail. An example of such a relation is ‘liking himself’. Since some people do not like themselves, this relation is not reflexive. Likewise, since some people do like themselves, it is not anti-reflexive. Like reflexivity, both symmetry and transitivity also have a middle category which neither has the property nor has the opposite of the property. As will be explained in the following sections, this middle category turns out to be very important for the philosophy of information.

Symmetry

A relation is symmetric if and only if

$$\forall a, b \in X, aRb \Rightarrow bRa$$

An example of a symmetric relation is ‘being a relative of’. If a is a relative of b , then b is also a relative of a . Similar to reflexivity, the opposite of symmetry (anti-symmetry) is not simply a failure of the original condition.

A relation is anti-symmetric if and only if

$$\forall a, b \in X, (aRb \wedge a \neq b) \Rightarrow \sim bRa$$

As the condition states, anti-symmetry requires that for no pair of the relation is its mirror image a member of the relation, unless the members of the pair are identical to each other. An example of an anti-symmetric relation is ‘being greater than’ as defined in the domain of numbers. If a is greater than b , then there is no way for b to be greater than a . As in the case of reflexivity, there are some relations that are neither symmetric nor anti-symmetric. ‘Being fond of someone’ is such a relation. If a is fond of b , b is also fond of a in some cases, but in other cases this may not be true. Thus, both symmetry and anti-symmetry conditions fail.

Transitivity

A relation is transitive if and only if

$$\forall a, b \in X, (aRb \wedge bRc) \Rightarrow aRc$$

A simple example of a transitive relation is identity. If a is identical to b and b is identical to c , then a has to be identical to c . As in the case of reflexivity and symmetry, the opposite of transitivity, which is called anti-transitivity, is not simply the failure of the transitivity condition. It states that the transitivity condition must not hold for any member of the relation. More formally:

A relation is anti-transitive if and only if

$$\forall a, b \in X, (aRb \wedge bRc) \Rightarrow \sim aRc$$

Although the common name for this characteristic is anti-transitivity, in order to be consistent with the philosophy of information literature, we shall refer to it here as intransitivity. An example of an intransitive relation is ‘being the mother of’. The transitivity condition fails for all pairs that fall under this relation. Whether or not there is a third category in which both transitivity and intransitivity fail is an important question. Some claim that preference relations (a is preferred

over b) are such an example. If a is preferred over b , and b is preferred over c , then a may be preferred over c in some cases and not in some others. One of the main claims of this chapter is that information-carrying relations are neither transitive nor intransitive.

Now we have completed our basic introduction to equivalence relations. In the following section, we will examine some selective proposals for identifying the content of information-carrying relations and then analyze those proposals in terms of the basic properties covered in this section.

INFORMATION-CARRYING RELATIONS: A BRIEF HISTORICAL SURVEY

When we say ‘information-carrying relations,’ what we intuitively mean by this is that some entity carries information about some other entity. This intuitive idea definitely points to a relation, but it is neither precise nor formal enough to be used in a theoretical framework. Step by step, it needs to be clarified and formalized. Let’s start with a simple information-carrying claim.

A carries information that B.

There are two important questions that need to be answered for clarifying this claim. First, what are A and B? In other words, what is the domain over which the relation is defined? The domain could be just propositions or it could also include natural signs. Let’s call this ‘the domain question.’ The second question is about the content of the ‘information-carrying relation.’ What does it mean to carry information about something else? To put it differently, how could we formalize the content of the relation? It is natural to call this ‘the content question.’

The Domain Question

Although Shannon’s mathematical work may be considered the starting point of the philosophy of

information, his work was mainly for engineering purposes. He clearly stated that his mathematical formalism does not deal with philosophical questions. When philosophers began using Shannon’s work for their own purposes, they labeled their efforts as ‘semantic theories of information’ in order to emphasize their interest in philosophical questions. For example, Bar-Hillel and Carnap (1952), in the earliest attempt of using information theoretic concepts in philosophy, called their theory ‘An Outline of a Theory of Semantic Information’. More or less, this trend has continued since then, and Floridi’s theory (2004), ‘A Strongly Semantic Theory of Information,’ is one of the most recent examples of this trend. Use of the qualifier ‘semantic’ is not just for emphasizing an interest in philosophical questions; it also gives us pointers as to the answer of the domain question. The word ‘semantic’ tells us that the members of the domain over which the information-carrying relations are defined must have an identifiable semantic content. Thus, the domain consists of propositions. Although restricting the domain to propositions is perfectly acceptable for some philosophical purposes, such a domain does not encompass all possible entities that may carry information. For example, natural signs such as smoke or dark clouds also carry information (Grice 1989). If that is the case, then the domain must include these signs, as well. Natural signs are not the only category that is an example of non-propositional information bearers. Some non-natural signs may also carry information. As Floridi puts it, “when you turned the ignition key, the red light of the low battery indicator flashed. This signal too can be interpreted as an instance of environmental information” (Floridi, in press). Although the red light is not a natural sign, it does still carry information. Floridi calls this type of information ‘environmental information.’ He accepts the legitimacy of such information but claims that it can be reduced to ‘semantic information’, which is necessarily propositional. Whether or not Floridi is right in his reductive claim is con-

troversial, but to pursue this question would take us too far away from the general framework of this chapter. Suffice it to mention that the jury is still out on Floridi's claim.

Dretske also focuses on propositional information in his theory of semantic information. He chooses all of his examples from propositions with identifiable content. However, he also acknowledges the possibility of information-bearers without an identifiable content, i.e., non-propositional information bearers. In his own words,

Up to this point examples have been carefully chosen so as to always yield an identifiable content. Not all signals, however, have informational content that lends itself so neatly and economically to propositional expressions. (Dretske 1981, p. 68)

Similar to Dretske and Floridi, Cohen and Meskin, in their exploration of a counterfactual theory of information, also use the set of propositions as the domain for information-carrying relations. Here is how they define information-carrying relations: "x's being F carries information about y's being G if the counterfactual conditional 'if y were not G, then x would not have been F' is non-vacuously true" (Cohen & Meskin 2006, p. 335). In this definition, the entities that may bear information clearly are propositions. However, this does not mean that Cohen and Meskin do not accept the possibility of non-propositional entities as information bearers; they only restrict their counterfactual analysis to propositions.

After this brief survey, we may conclude that almost all theories of semantic information identify the domain as the set of propositions. However, non-propositional entities such as natural signs also need to be taken into account while identifying the proper domain for information-carrying relations. Thus, the conclusion is that the fundamental properties of information-carrying relations must be analyzed for two different possible domains: one that includes only propositions and another

that includes non-propositional signals as well as propositions.

The Content Question

To identify the domain of a relation is the first order of business, but it is not the whole story. The content of a relation also needs to be unambiguously determined. For most relations, this task is straightforward. For example, 'the greater than' relation for numbers is clear and unambiguous, and so is the 'being the father of' relation. Any controversy about whether this relation holds between two human beings can be resolved with a DNA test. In the case of information-carrying relations, however, the situation is rather messy because the concept of information is prevalently used in many different senses (Floridi, in press; Allo 2007; Scarantino and Piccinini, in press). In fact, whether or not there could be a common denominator for all different uses of information is also not clear. Shannon himself pointed out this wide usage of information and also stated his suspicion about the existence of a common denominator.

The word 'information' has been given different meanings by various writers in the general field of information theory. It is likely that at least a number of these will prove to be useful in certain applications to deserve further study and permanent recognition. It is hardly to be expected that a single concept of information would satisfactorily account for the numerous possible applications of the general field. (1950, p. 80)

Given this prevalent and ambiguous usage of the notion of information, philosophers who need to identify the content of such relations can only proceed by providing a formalism for the meaning of 'information' that they use. There have been several attempts to do so. A brief historical survey, starting with Shannon's formalism, is useful for

understanding the evolution of these endeavors. Before we start our historical survey, though, a disclaimer is in order. This is by no means a complete historical survey; rather, I included four representative theories. Bar-Hillel and Carnap's theory is included because it is the first example of a semantic theory of information. Dretske's theory is covered because of its scope and influence. Harms' and Cohen and Meskin's theories are surveyed because they represent different attitudes toward the relevance of Shannon's mathematical theory of information. There are several important philosophical theories of information that had to be left out due to the space constraints of this chapter. Some examples are Sayre (1976), Devlin (1991), Barwise and Seligmann (1997) and Floridi (2004). Needless to say, exclusion from this chapter does not represent any judgment about the quality of those semantic theories of information.

Since the first edition of Shannon's seminal article, "The Mathematical Theory of Communication", both philosophers and psychologists began adopting the notions that Shannon develops for their own purposes. They realized the potential value of notions such as information, entropy and channels for solving philosophical and psychological problems. After all, the relation between the human mind and the external world is one of communication, and Shannon's formalism has proven to have a high explanatory power for the notions of communication channels and information transmission. After that initial enthusiastic reaction, however, philosophers realized that there are fundamental differences between Shannon's information and the notion that they need for their own philosophical purposes. Shannon's goal was to formalize the best method for coding and encoding messages for communication purposes. Given these engineering purposes, he had to work at an abstract level at which the content of a signal did not matter. After all, he needed a theory that could be applied to any content that might be communicated.

Shannon's main question was the following: given a set of possible states, what is the expected surprisal value of a particular state that belongs to the set of all possible states? More formally, he strove to determine the expected value of a random r_i where r_i is a member of $S = \{r_1, r_2, r_3, \dots, r_n\}$. He started out with three basic intuitions:

- i. The expected value should depend only on the probability of r_i , not on the content of r_i ;
- ii. The expected surprise should be a kind of expected value;
- iii. The expected surprisal value of an r_i should increase as the r_i 's become more equiprobable.

The last intuition is similar to the case of a fair and unfair coin. The result of a toss of a fair coin is less anticipated than that of an unfair coin. Surprisingly enough, the only set of functions that satisfy these three intuitions is the set of entropy functions of thermodynamics.² The very first of Shannon's three basic intuitions implies that his theory is not about the content of a signal, but rather it is about the amount of information that a signal or a probability distribution for a set of states has. Shannon clearly stated this fact in their seminal work: "These semantic aspects of communication are irrelevant to the engineering problem" (Shannon 1948, p. 382). In a similar vein, Colin Cherry, another communication engineer, emphasized this aspect of the mathematical theory of communication: "It is important to emphasize, at the start, that we are not concerned with the [content] or the truth of messages; semantics lies outside of the scope of mathematical information theory" (Cherry 1951). This statement about the mathematical theory of information shows the point where Shannon's and the philosophers' interests diverge. Philosophers are interested in identifying the content of a signal, whereby the signal may be a linguistic or mental entity. This

divergence led philosophers to search for a more suitable notion of information. Bar-Hillel and Carnap's theory of semantic information was the earliest such attempt. Their theory is based on Carnap's logical analysis of probability (Carnap 1950). Accordingly, the content of an informational signal can be defined negatively by the set of all possible state descriptions that are excluded by the signal (Floridi, in press, p. 141). Although this was a step toward identifying the content of an informational signal consistent with the requirements of Shannon's mathematical theory of information, Bar-Hillel and Carnap's theory had a serious shortcoming that leads to a paradoxical situation. In their formalization, contradictions carry an infinite amount of information (Bar-Hillel & Carnap 1952, p. 21).³ This feature of their theory, to say the least, is implausible.

Dretske, in his 1980 book *Knowledge and the Flow of Information*, also tried to provide a semantic theory of information consistent with the mathematical theory of information. Dretske's framework deserves attention for three main reasons. First, Dretske attempted to explain perceptual content, belief and knowledge in terms of informational content. In that respect, his framework encompasses a wide range of philosophical issues. Second, his theory is the earliest example of one of the relational properties of information carrying, (i.e., transitivity, playing a central role). Despite its importance, however, transitivity also leads to a controversial feature of the theory. For Dretske, information necessarily implies truth. In other words, propositions that are not true do not carry information. Third, according to Dretske, Shannon's mathematical theory of information is not very useful for epistemology or philosophy of mind because it is about the average information that a set of messages contains, whereas epistemology and philosophy of mind are concerned about whether a person knows or acquires a particular fact on the basis of a particular signal. In other words, philosophical issues hinge on the specific content of information, not just the amount of

information that a signal carries. Despite these diverging interests, Dretske believes that some notions of Shannon's theory could be a starting point for solving philosophical problems. He borrows the notion of entropy of a signal from Shannon and develops his semantic theory of information.

Several philosophers, however, questioned some of Dretske's claims. The first targeted claim was the inseparable connection between information and truth. Some found truth encapsulation, i.e., that if a signal carries information about p , then p has to be true, to be too demanding. The second target was Dretske's claim about the lack of usefulness of Shannon's mathematical theory. Contrary to Dretske, several philosophers claimed that Shannon's theory could be more useful for philosophical purposes. For example, Grandy (1987) provided an information theoretic approach based on Shannon's mutual information, and claimed that a proper use of mutual information could serve as a basis for an ecological and naturalized epistemology. Similarly, Harms (1998) claimed that mutual information provides an appropriate measure of tracking efficiency for the naturalistic epistemologist, and that this measure of epistemic success is independent of semantic maps and payoff structures. Usher (2001) proposed a naturalistic schema of primitive conceptual representations using the statistical measure of mutual information. In order to see how the notion of mutual information develops in regard to philosophical problems, it is useful to examine this line of reasoning, but because of current lack of space, it is not possible to thoroughly cover these three attempts. We shall therefore pick Harms' framework as the representative of this line of reasoning and examine it in detail in the next section.

Finally, there is one more theory that needs to be included in our historical survey and analysis: Cohen and Meskin's counterfactual theory of information. The importance of the counterfactual theory of information lies in the fact that it does not borrow any notions from Shannon's math-

ematical theory of information. In fact, Cohen and Meskin claim that using Shannon's insight about entropy and uncertainty reduction, as in Dretske's framework, leads to a doxastic theory of information and not an objective one (Cohen & Meskin 2006, p. 340). Moreover, their theory is the first in the literature in which the transitivity of information-carrying relations is questioned. Thus, in terms of their take on Shannon's theory and transitivity, they represent a line of reasoning different from both Dretske's and Harms'.

To review, in this section we have surveyed three suggestions for answering the 'content question.' Dretske identifies the content of information-carrying relations in terms of entropy and conditional probabilities. In that respect, his framework utilizes some conceptual tools of Shannon's mathematical theory of information, leading to an independent semantic theory in which perceptual content, belief and knowledge are accounted for in terms of informational content. Contrary to Dretske, Harms thinks that Shannon's framework is much more in line with semantic purposes of philosophers, and he develops a theory based on the notion of mutual information. Cohen and Meskin think that the fundamental insights and concepts of Shannon's mathematical theory are not good candidates for clarifying semantic issues; instead, they suggest a counterfactual account. Thus, an analysis of these three theories in terms of relational properties (reflexivity, symmetry and transitivity) will cover the landscape of the literature of philosophy of information to a satisfactory extent. Doing so is the main task of the next section.

THREE DEFINITIONS OF SEMANTIC CONTENT

In this section, we will analyze Dretske's, Harms' and Cohen and Meskin's theories in more detail and also evaluate their suggestions for identifying

informational content in terms of the fundamental relational properties.

Dretske

Dretske bases his theory on the notion of informational content. By using this notion, together with the tools of Shannon's theory, Dretske aims to give an account of mental content, perception, belief and knowledge. Dretske defines the notion of informational content as follows:

Informational Content: *A signal r carries the information that s is $F =$ the conditional probability of s 's being F , given r (and k), is 1 (but, given k alone, less than 1) [k refers to background knowledge]. (Dretske 1981, p. 65)*

As a result of assigning unity to the conditional probability, Dretske rejects the possibility of partial information and misinformation. He says that "information is certain; if not, it is not information at all." Although this claim, dubbed the 'Veridicality Thesis' by Floridi, is useful for some philosophical purposes (e.g., semantic analysis of true propositions; Floridi 2007), for some other purposes (e.g., accounting for mental representation; Demir 2006) it turns out to be counter-productive. Before making any judgment about this issue, it is important to understand Dretske's rationale for insisting on this claim about information-carrying relations. His main rationale is to distinguish genuine information-carrying relations from coincidental correlations. If your room and my room have the same temperature at a given moment, the thermometers in both rooms will show the same temperature, yet it would be wrong to say that the thermometer in your room carries information about my room's temperature. For an information-carrying relation, there needs to be some lawful dependency between the number that the thermometer shows and the temperature of the room. This dependency holds between the

thermometer in my room and my room's temperature. There is no such dependency between my room's temperature and the thermometer in your room. The nomic dependency requirement does not directly appear in Dretske's informational content definition. However, assigning unity to the conditional probability in the definition is a direct result of nomic dependencies.

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 occurrence when s is not F . (Dretske 1981, p. 245, emphasis original)

Besides this rationale, Dretske presents three arguments for claiming that the value of conditional probability in his definition of informational content must be 1, nothing less. Although many scholars have questioned the legitimacy of assigning unity to conditional probabilities, Dretske says that no one has attempted to reject his arguments (Dretske 1983, p. 84-85).⁴ What Dretske says is true. There is no comprehensive attempt to reject his arguments. In this chapter, we will be focusing only on one of them, the argument from transitivity, and will only briefly mention the other two.

Dretske's first argument rests on the transitivity of information-carrying relations. He claims that information flow is possible only if the flow is transitive, (i.e., if a signal A carries the information B , and if B carries the information C , then A must also carry the information C). He calls this property of transitivity the Xerox Principle. The name is straightforward. The photocopy of a photocopy of a document has the same printed information as the original. The only way of accommodating this principle within his conditional probability framework is to assign unity, because it is a simple mathematical fact that conditional probabilities are not transitive unless they are equal to 1. Hence, in order to satisfy the transitivity property, (i.e., the Xerox Principle), conditional probabilities must be 1.

Secondly, Dretske says that "there is no arbitrary place to put the threshold that will retain the intimate tie we all intuitively feel between knowledge and information." If the information of 's's being F ' can be acquired from a signal which makes the conditional probability of this situation happening something less than 1—say, for example 0.95—then "information loses its cognitive punch" (Dretske 1981, p. 63).

The principle that he uses for his third argument is a close relative of the Xerox Principle, and he calls it the Conjunction Principle. If a signal carries the information that B has a probability of p_1 and the information that C has a probability of p_2 , the probability of carrying the information that B and C must not be less than the lowest of p_1 and p_2 . However, again it is a simple mathematical fact that this could not happen with conditional probabilities if they are less than 1.

Dretske's arguments become more intuitive once thought of as a result of a learning metaphor. For Dretske, information-carrying relations are very similar to, if not identical to, 'learning' relations. If I can learn B from A and C from B , then I should be able to learn C from A . This intuitive claim is nothing but the Xerox Principle, i.e., transitivity. 'Learning B from A ' is identical to ' A carries the information that B .' Likewise, 'learning C from B ' means ' B carries the information that C .' These two together imply that 'I can learn C from A ,' i.e., A carries the information that C . A similar reasoning applies to the Conjunction Principle. For the Arbitrary Threshold Thesis, since the metaphor is to learn, we ideally want to learn the truth, not an approximation of truth. In short, Dretske's intuitive motivation for his arguments is the metaphor of learning.

After this brief exposition, we are now in a position to evaluate the fundamental properties of Dretske's informational content. His definition is reflexive because the conditional property of a signal's content (s is F), given the same signal (s is F), is 1:

Table 1.

	Reflexive	Anti-Reflexive	Symmetric	Anti-Symmetric	Transitive	Intransitive
<i>Dretske's Theory</i>	√	X	X	X	√	X

$$\Pr ('s \text{ is } F' | 's \text{ is } F') = 1$$

Since the definition is reflexive, we automatically know that it is not anti-reflexive. Dretske's definition is neither symmetric nor anti-symmetric. For symmetry to hold, for any signal, such as '*s is F*,' if it carries information that '*t is G*,' then '*t is G*' must also carry the information that '*s is F*.' The antecedent of this conditional implies the following equation:

$$\Pr ('s \text{ is } F' | 't \text{ is } G') = 1$$

This equation, however, does not guarantee that the conditional probability of '*t is G*' given '*s is F*' is 1, which is required for the truth of the consequent of the symmetry conditional. In some cases, $\Pr ('t \text{ is } G' | 's \text{ is } F')$ will be less than 1; in others, it might be exactly 1. Thus, Dretske's definition is neither symmetric nor anti-symmetric.

Lastly, it is obvious that Dretske's definition is transitive. He builds his very framework on the basis of transitivity. We can summarize these findings with a simple table (see Table 1)

Harms

Harms (1998), in "The Use of Information Theory in Epistemology," undertakes an ambitious task which has two main components:

- i. To identify the relevant measure of information for tracking efficiency of organisms;
- ii. To flesh out the relationship between the information measure and payoff structures.

For the first part of his task, he offers mutual information as the right tracking efficiency measure. For the second part, he shows that mutual information is independent of payoff structures. Although both of these tasks are philosophically important, for our immediate purposes we shall focus only on the first one.

For identifying the relevant measure of efficiency tracking, Harms looks to Shannon's mathematical theory of information because he thinks that the gap between Shannon's theory and philosophically relevant aspects of information is not as large as Dretske and some others think. As previously mentioned, Shannon's theory focuses on measuring the amount of information that a signal carries, whereas philosophers are interested in the content of the signal, not just the amount. As a result, some philosophers think of Shannon's theory as not very relevant to philosophical questions. Harms disagrees:

It is one thing to calculate the accuracy of sending and receiving signals; it is another thing entirely to say what those messages are about, or what it means to understand them. Consequently, one might think that since the notion is not semantic, it must be syntactic or structural. The dichotomy is false, however. What communication theory offers is a concept of information founded on a probabilistic measure of uncertainty. However, even respecting that information theory does not presume to quantify or explain meaning, there remains the possibility that the information theoretic notion of information can be applied to semantic problems. (Harms 1998, p. 481)

The Fundamental Properties of Information-Carrying Relations

The notion of mutual information, which Harms offers as a good candidate for philosophical purposes, is simply a similarity measure between two variables. It has properties that are useful for defining semantic content of informational signals or messages. In Shannon's formalism, the mutual information between two variables, $I(s,r)$, is defined as follows. Lower case letters without subscripts are random variables and lower case letters with subscripts are the values of random variables (please see the glossary for the definition of random variables).

$$\text{(Mutual Information)} I(s,r) = - \sum \Pr(s_i) \log \Pr(s_i) + \sum \Pr(s_i | r_j) \log \Pr(s_i | r_j)$$

Since the probability values range between 0 and 1, the logarithm of a probability is always negative, and as a result, the second term in the above equation is negative. But the first term, because of the minus sign, is positive. Thus, the highest value of the mutual information between two variables is the same as the value of the first term. The first term is nothing other than the entropy of the first variable (again, please see the glossary). Hence, the amount of mutual information ranges between zero and the entropy of the first variable. This leads to an interesting result for information-carrying relations: whether or not A carries the information that B is not a "yes" or "no" issue anymore. In Dretske's framework, two signals either enter into an information-carrying relation or they do not; there is no gradation between these two options. The situation is different in Harms' theory; information-carrying relations lie on a continuum in his framework.

This brief exposition of Harms' theory provides enough ground for evaluating his definition's properties. For reflexivity, we need to calculate the amount of mutual information of a signal with itself.

$$\text{(Reflexivity)} I(s,s) = - \sum \Pr(s_i) \log \Pr(s_i) + \sum \Pr(s_i | s_i) \log \Pr(s_i | s_i)$$

Since $\Pr(s_i | s_i)$ is 1 and the logarithm of 1 is 0, the second term is 0. $I(s,s)$ ends up being equal to the first term of the equation, which is the entropy of the variable s . As stated above, this is the highest possible value of mutual information. In other words, a signal has the highest amount of mutual information with itself. Thus, mutual information is reflexive.

For symmetry to hold, the following two equations need to return the same value.

$$\text{(Symmetry 1)} I(\mathbf{s},\mathbf{r}) = - \sum \Pr(s_i) \log \Pr(s_i) + \sum \Pr(s_i | r_j) \log \Pr(s_i | r_j)$$

$$\text{(Symmetry 2)} I(\mathbf{r},\mathbf{s}) = - \sum \Pr(r_j) \log \Pr(r_j) + \sum \Pr(r_j | s_i) \log \Pr(r_j | s_i)$$

Since it only takes basic knowledge of algebra and probability to show that these equations are equal to each other, we leave the proof to the interested reader and conclude that mutual information is a symmetric notion.

Whether or not mutual information is transitive is a bit more complicated than determining symmetry and reflexivity. Under special circumstances, it turns out to be transitive. But in most cases, it is not transitive. This fact is a corollary of the Data Processing Inequality theorem of the mathematical theory of communication.

Data Processing Inequality Theorem: If there is an information flow from A to C through B , then the mutual information between A and B is greater than or equal to the mutual information between A and C . More formally:

$$\text{If } A \rightarrow B \rightarrow C \text{ then } I(A,B) \geq I(A,C). \text{ (Cover 1991, p. 32)}^5$$

For transitivity to hold for mutual information, if A carries information that B , and B carries information that C , then A has to carry information that C . In other words, the mutual information between A and B needs to be equal to the mutual information between A and C . As the theorem

Table 2.

	Reflexive	Anti-Reflexive	Symmetric	Anti-Symmetric	Transitive	Intransitive
<i>Dretske's Theory</i>	√	X	X	X	√	X
<i>Harms' Theory</i>	√	X	√	X	X	X

suggests, equality happens only in some cases; in other cases, $I(A, C)$ turns out to be smaller than $I(A, B)$. Thus, mutual information is transitive only in some cases, but not in others.

Now, let's expand our summary table (see Table 2) from the previous section with our new findings.

Cohen and Meskin

Cohen and Meskin, in their 2006 article, "An Objective Counterfactual Theory of Information," explore an alternative route for defining informational content. Their motivation for seeking an alternative is to avoid using the notion of conditional probabilities. As we saw in the previous sections, both Dretske and Harms appeal to conditional probabilities in their definitions of informational content. In fact, this has been the standard approach since Shannon. Appeals to conditional probabilities, however, come with many problems related to the notion of probability and its interpretations.⁶ Cohen and Meskin suggest a radically different alternative that appeals to counterfactuals instead of probabilities.

In their paper, Cohen and Meskin begin with a crude version of their counterfactual theory, and then revise it by adding a non-vacuousness clause to avoid some difficulties concerning necessary truths. For both the crude and revised accounts, they present one weak and one strong version. The weak versions take the counterfactual criterion as only a sufficient condition for information-carrying relations, whereas the strong versions take it as both necessary and sufficient. The difference between their strong and weak versions

is irrelevant for our purposes. Hence, for the sake of simplicity, we shall state only the weak version of their claim:

x's being F carries information about y's being G if the counterfactual conditional 'if y were not G, then x would not have been F' is non-vacuously true. (Cohen & Meskin 2006, p. 335)

The non-vacuousness clause excludes assigning the information-carrying relation to cases where *y's being G* is necessarily true. If *y's being G* is necessarily true, then the counterfactual will prove to be true no matter what; hence, the counterfactual will be vacuously true. Following the generally accepted intuition that necessary truths carry no information at all,⁷ Cohen and Meskin aim to exclude necessary truths from the set of information-carrying signals by adding the non-vacuousness clause. Cohen and Meskin argue that the counterfactual theory of information may be preferable to the standard approaches. Leaving aside the issue of whether or not their claim is true, we shall proceed to analyze the properties of their counterfactual definition.

Since the information-carrying relation is determined by a conditional in Cohen and Meskin's framework, the relation automatically becomes reflexive. Any conditional, for which the antecedent and the consequent are the same propositions, is always true.

For symmetry, we have to assume that one signal, say, *x's being F*, carries information about another signal, say, *y's being G*. This assumption leads to the truth of the following counterfactual:

The Fundamental Properties of Information-Carrying Relations

(Counterfactual 1) 'If y were not G , then x would not have been F ' is non-vacuously true.'

If this conditional implies its converse, then symmetry holds for Cohen and Meskin's counterfactual definition; otherwise, it does not. The converse claim is the following:

(Counterfactual 2) 'If x were not F , then y would not have been G ' is non-vacuously true.'

The standard semantics for evaluating the truth condition of counterfactuals is Lewis' possible worlds (Lewis 1973; also see the Glossary). In Lewis' semantics, the first counterfactual is true if and only if in the closest world where y is not G , x is not F . Let's call this world w_1 . For the second counterfactual to be true, in the closest world where x is not F , y must be not G . Let's call this world w_2 . The truth condition of the first counterfactual does not guarantee the truth condition of the second counterfactual, because there could be a world closer than w_2 where x is not F and yet y is G . In some cases, by coincidence, w_2 might turn out to be the closest to the actual world, but this is not necessarily the case. Thus, the counterfactual definition is neither symmetric nor anti-symmetric.

It is a well-established fact that counterfactual conditionals are not transitive.⁸ The simplest way of seeing this fact is to evaluate the validity of the following inference schema:

- A counterfactually implies B.
- B counterfactually implies C.

Therefore,

- A counterfactually implies C.

This inference is NOT valid because the closest possible A-world may not be a C-world, given that the closest possible A-world is a B-world and the closest possible B-world is a C-world. So even if the conclusion follows from the premises in some cases, there could be other cases in which it does not. Thus, the counterfactual definition of information-carrying relations is neither transitive nor intransitive. In Cohen and Meskin's own words,

[The counterfactual] account implies that the information-carrying relation is non-transitive, it does not imply that the information-carrying relation is intransitive. Our account denies that information-carrying is transitive tout court, but it allows that in many (but not all) cases information may flow from one event to another along a chain of communication. (Cohen & Meskin 2006, p. 340)

We could complete our summary table (see Table 3) by adding the results of the analysis of Cohen and Meskin's counterfactual theory.

As the above table suggests, there is a consensus about reflexivity. This is only to be expected because, after all, a signal, if it has some informational content, will carry information about its own content. The disagreement, however, arises in the cases of symmetry and transitivity. The next section focuses on this disagreement.

Table 3.

	Reflexive	Anti-Reflexive	Symmetric	Anti-Symmetric	Transitive	Intransitive
<i>Dretske's Theory</i>	√	X	X	X	√	X
<i>Harms' Theory</i>	√	X	√	X	X	X
<i>C & M's Theory</i>	√	X	X	X	X	X

INFORMATION-CARRYING RELATIONS: A GENERAL ANALYSIS

Among the three theories analyzed above, only Harms' theory is symmetric. Symmetry, although it provides neat mathematical features, may not be a desirable feature for some philosophical purposes. One of the long-standing projects in philosophy of mind is to give a naturalistic account of mental representation. The notion of information seems to be a promising candidate for the foundation of such a naturalistic account,⁹ because, after all, a mental state acquires information from a state of affairs in the world, and thus carries information about that state of affairs. If this simple intuition is right, however, the required information-carrying relation needs to be not symmetric. My mental state that represents a dog carries information about the dog in my yard, but the dog in my yard does not carry information about my mental state. Of course, the dog in the yard causes my mental state, but it would be wrong to claim that it carries information about my mental state. In other words, information-carrying relations are different than causal relations. For the goal at hand, i.e., to account for mental representation, any symmetric notion of informational content fails to do the job. A similar story could easily be told for linguistic representation as well. Thus, ideally, we want a non-symmetric conceptualization of information-carrying relations, especially for explaining mental and linguistic representation.

After this short analysis, one may conclude that Harms' claim about the relevance of Shannon's theory for philosophical purposes is wrong (please see the quotation in Section 3.2.). However, this would be too quick of a judgment, because there may be some other notions within the rich repertoire of the mathematical theory of information that may serve better for Harms' theory. In fact, there is a very good candidate that has all the desired features of mutual information without being symmetric: it is the Kullback-Leibler divergence

measure. Further research is needed for evaluating the plausibility of this measure.

For assessing whether or not transitivity is a desired feature for information-carrying relations, the Data Processing Inequality theorem, as stated in Section 3.2., is crucial. For the ease of readers, let us state the theorem once again:

Data Processing Inequality Theorem: *If $A \rightarrow B \rightarrow C$ then $I(A,B) \geq I(A,C)$*

Transitivity holds only for the equality condition in the greater than or equal to relation between $I(A,B)$ and $I(A,C)$. For other cases, transitivity fails. The equality condition occurs only if the chain formed by the information from A to C through B ($A \rightarrow B \rightarrow C$) is a Markov chain¹⁰. A Markov chain occurs when the conditional distribution of C depends only on B and is independent of A . Obviously, this is a very strict constraint, and it is rarely true in real life information channels. If this constraint is not fulfilled, then the probability of having equality becomes lower and lower as the chain of information flow becomes longer. Hence, transitivity is valid only in idealized cases. Once again, it should be noted that transitivity corresponds to the equality between $I(A,B)$ and $I(A,C)$ in the data processing equality theorem, not the greater than relation.

Markov chains, i.e., informational chains where only the two subsequent members of the chain conditionally depend on each other, are not strong enough to exploit the statistical regularities that may exist in an informational source. Shannon, in his seminal article, "The Mathematical Theory of Communication", showed the importance of longer conditional dependencies in a sequence for exploiting the statistical regularities in an informational source (Shannon 1948, p.413-416). The informational source that he chose was English. As it is known, some letters are more frequent than others in English. This is the main reason for assigning the highest point value to the letter

The Fundamental Properties of Information-Carrying Relations

Table 4. Improvement index

	Meaningful Sequences (MS)	Length of MS	Total Length	Success Index	% Increase
1st Order	-	0	72	0	NA
2nd Order	ON, ARE, BE, AT, ANDY	13	118	0.11	NA
3rd Order	IN, NO, IN NO, WHEY, OF, OF, THE, OF THE, IS OF	30	108	0.28	154%

Q in Scrabble; it is the least frequently used letter in the English language. This is an important statistical regularity of English, but not the only one. There are also patterns depending on the previous letters that occur in a sequence. For example, the probability of having an ‘S’ after an ‘I’ is different than the probability of having a ‘C’ after an ‘I.’ Similarly, the probability of having a ‘U’ after the sequence ‘YO’ is different than the probability of having an ‘R.’ Shannon used these statistical patterns in sequence in order to produce intelligible sequences in English without feeding any extra rule to the sequence-producing mechanism. For all sequences, he assumed a 27-symbol alphabet, the 26 letters and a blank. In the first sequence, he used only the occurrence frequencies of letters; he called this “first-order approximation.” The idea behind the process by which he produced the sequence can be thought of in the following way. Imagine a 27-sided die upon which each side is biased according to its occurrence frequency. Then, by simply rolling the die at each step, one decides the symbol that should appear for that step. The output of his first sequence where only letter frequencies are used is the following:

- **First-Order Approximation**
OCRO HLI RGWR NMIELWIS
EU LL NBNESEBYA TH EEI AL-
HENHTTPA OOBTTVA NAH BRL.

For the second sequence, the frequencies that he used were the frequency of a letter given the letter that comes just before E. That is to say, instead of using the simple occurrence frequency

of the letter E, he used the conditional frequency of E given the previous letter. For example, if the previous letter were K, then he used the occurrence frequency of E given K. This is his second-order approximation.

- **Second-Order Approximation**
ON IE ANTSOUTINYS ARE T INC-
TORE ST BE S DEAMY ACHIN D
ILONASIVE TUCCOOWE AT TEA-
SONARE FUSOTIZINANDYTOBE
SEACE CTISBE.

In the third-order approximation, he used the occurrence frequencies of letters given the previous two letters instead of one.

- **Third-Order Approximation**
IN NO IST LAT WHEY CRATICT
FROURE BIRS GROCID PONDE-
NOME OF DEMONSTURES OF
THEREPTAGINISREGOACTIONA
OF CRE.

There is an improvement from the second-order approximation to the third-order. This improvement may not seem significant at first glance. However, when measured quantitatively, the third-order approximation almost triples the success of the second-order approximation. Unfortunately, Shannon did not provide such a quantitative success index, because for his purposes the improvement was noticeable enough. A simple success index that can be used is the ratio of the length of the meaningful sequence to the

Table 5.

	Reflexive	Anti-Reflexive	Symmetric	Anti-Symmetric	Transitive	Intransitive
<i>Dretske's Theory</i>	√	X	X	X	√	X
<i>Harms' Theory</i>	√	X	√	X	X	X
<i>C & M's Theory</i>	√	X	X	X	X	X
<i>Ideal Expectation</i>	√	X	X	X	X	X

length of the entire sequence. The success index values calculated accordingly are shown in Table 4. The index value of the third-order approximation is equal to 2 ½ times the index value of the second-order approximation. That is to say, there is a significant improvement from the second to the third order, and the level of improvement increases exponentially when one moves to higher order approximations such as the fourth-order, the fifth-order and so on.

In short, the sequence of English letters becomes much more meaningful when one increases the length of the dependencies in conditional probabilities. In other words, a successful use of statistical regularities requires longer informational chains in which the conditional probability of an entity depends not just on the previously occurring one, but on several others that come before that entity. Shannon's second-order approximation, conditional probabilities given just the previous letter, corresponds to the idea of Markov chains as mentioned above. Dretske's insistence on transitivity presumes a Markov chain and hence stops at the second-order approximation level. However, the amount of information that one can exploit from an informational source by a Markov chain is very limited, as shown in Shannon's second-order approximation. Most of the informational sources (for example, natural languages and the external world) are much richer, and to exploit such richness one needs to extend dependencies beyond the limits of a Markov chain. As Table 4 shows, even going one order level up from a strict Markov chain significantly increases the ability

to exploit regularities in an informational source. Hence, transitivity is not a desirable feature for such purposes.

In this section, we have concluded that a non-symmetric and non-transitive approach to identifying the content of information-carrying relations will serve better for some philosophical purposes. This means that an information-carrying relation is not an equivalence relation. Although equivalence is needed for a neatly organized domain of informational entities, it turns out that reality is much messier than we would like it to be. Let's add these findings into our summary table (see Table) for a complete visual depiction.

CONCLUSION

In this chapter, we have completed a comprehensive analysis of information-carrying relations in terms of fundamental mathematical properties: reflexivity, symmetry and transitivity. As Table 5 above depicts, a reflexive, non-symmetric and non-transitive content definition is better suited for philosophical purposes. Given this, it looks as though Cohen and Meskin's counterfactual theory of semantic information is the best available candidate for the philosopher's ideal expectation. This result, however, needs to be taken with a grain of salt, because Cohen and Meskin's theory completely avoids Shannon's formalism. Shannon's mathematical theory of information has proven to have a high explanatory power for the technical features of information flow. The

main motivation for using information-theoretic notions for solving philosophical problems was to exploit this explanatory power. Cohen and Meskin's counterfactual theory does not have this benefit because it avoids Shannon's formalism. Whether or not this is a price worth paying is an important question that requires further research.

Although the chapter provides a thorough analysis of the issue at hand, it does, by necessity, leave some questions unanswered. Attempting to answer these questions will be an essential part of the future trends in the literature. For now, let us state three of these questions as suggestions for future research:

- i. What is the role of non-propositional information bearers for the philosophically relevant analysis of information flow?
- ii. Is there a non-symmetric notion within the repertoires of the mathematical theory of information that successfully accounts for information measure and payoff structures, as Harms' theory does? As suggested above, the Kullback-Leibler divergence measure seems to be a good candidate for this purpose and it needs to be analyzed from this perspective.
- iii. Is it possible to provide a probabilistic version of Dretske's informational content? A probabilistic version of Dretske's definition would carry the exact definition without assigning unity to conditional probabilities. In this way, some of the seemingly implausible consequences of assigning unity would be avoided.

'If A had been the case, C would have been the case' is true (at a world w) iff (1) there are no possible A-worlds (in which case it is vacuous), or (2) some A-world where C holds is closer (to w) than is any A-world C does not hold (Lewis 1973, p. 560).

REFERENCES

- Allo, P. (2007). Logical pluralism and semantic information. *Journal of Philosophical Logic*, 36(4), 659–694. doi:10.1007/s10992-007-9054-2
- Bar-Hillel, Y., & Carnap, R. (1952). *An outline of a theory semantic information* (Tech. Rep. No.247). Cambridge, MA: Massachusetts Institute of Technology, Research Laboratory of Electronics.
- Barwise, J., & Seligman, J. (1997). *Information Flow: The Logic of Distributed Systems*. Cambridge, UK: Cambridge University Press.
- Bremer, M. E. (2003). Do Logical Truths Carry Information? *Minds and Machines*, 13, 567–575. doi:10.1023/A:1026256918837
- Brogaard, B., & Salerno, J. (2008). Counterfactuals and context. *Analysis*, 68, 39–45. doi:10.1093/analys/68.1.39
- Carnap, R. (1950). *Logical Foundations of Probability*. Chicago, IL: University of Chicago Press.
- Cherry, C. E. (1951). A History of the Theory of Information. In *Proceedings of the Institute of Electrical Engineers*, 98.
- Cohen, J., & Meskin, A. (2006). An Objective Counterfactual Theory of Information. *Australasian Journal of Philosophy*, 84, 333–352. doi:10.1080/00048400600895821
- Demir, H. (2006). *Error Comes with Imagination: A Probabilistic Theory of Mental Content*. Unpublished doctoral dissertation, Indiana University, Bloomington, Indiana.
- Demir, H. (2008). Counterfactuals vs. conditional probabilities: A critical analysis of the counterfactual theory of information. *Australasian Journal of Philosophy*, 86(1), 45–60. doi:10.1080/00048400701846541
- Devlin, K. J. (1991). *Logic and Information*. Cambridge, UK: Cambridge University Press.

- Dietrich, E. (2007). Representation . In Thagard, P. (Ed.), *Handbook of Philosophy of Science: Philosophy of Psychology and Cognitive Science* (pp. 1–30). Amsterdam: Elsevier. doi:10.1016/B978-044451540-7/50018-9
- Dretske, F. (1981). *Knowledge and the flow of information*. Cambridge, MA: MIT Press.
- Dretske, F. (1983). Precis of Knowledge and the flow of the information. *The Behavioral and Brain Sciences*, 6, 55–63. doi:10.1017/S0140525X00014631
- Floridi, L. (2004). Outline of a theory of strongly semantic information. *Minds and Machines*, 14(2), 197–222. doi:10.1023/B:MIND.0000021684.50925.c9
- Floridi, L. (in press). *The Philosophy of Information*. Oxford, UK: Oxford University Press.
- Grandy, R. E. (1987). Information-based epistemology, ecological epistemology and epistemology naturalized. *Synthese*, 70, 191–203. doi:10.1007/BF00413935
- Grice, P. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.
- Harms, W. F. (1998). The use of information theory in epistemology. *Philosophy of Science*, 65, 472–501. doi:10.1086/392657
- Hintikka, J. (1970). Surface information and depth information . In Hintikka, J., & Suppes, P. (Eds.), *Information and Inference* (pp. 263–297). Dordrecht, The Netherlands: Reidel.
- Kyburg, H. E. (1983). Knowledge and the absolute. *The Behavioral and Brain Sciences*, 6, 72–73. doi:10.1017/S0140525X00014758
- Lehrer, K., & Cohen, S. (1983). Dretske on Knowledge. *The Behavioral and Brain Sciences*, 6, 73–74. doi:10.1017/S0140525X0001476X
- Lewis, D. (1973). *Counterfactuals*. Oxford: Blackwell.
- Loewer, B. (1983). Information and belief. *The Behavioral and Brain Sciences*, 6, 75–76. doi:10.1017/S0140525X00014783
- Sayre, K. M. (1976). *Cybernetics and the Philosophy of Mind*. London: Routledge & Kegan Paul.
- Scarantino, A., & Piccinini, G. (in press). Information without Truth. *Metaphilosophy*.
- Shannon, C. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27, 379-423 & 623-656.
- Shannon, C. E. (1993). The Lattice Theory of Information . In Sloane, N. J. A., & Wyner, A. D. (Eds.), *Collected Papers*. Los Alamos, CA: IEEE Computer Society Press.
- Usher, M. (2001). A statistical referential theory of content: Using information theory to account for misrepresentation. *Mind & Language*, 16, 311–334. doi:10.1111/1468-0017.00172

KEY TERMS AND DEFINITIONS

Counterfactuals: A counterfactual is a conditional where the antecedent is a non-factual statement. For material conditionals, when the antecedent is false, then the conditional is automatically true. However, this is not the case for counterfactual conditionals as can be seen from the following example: If Oswald had not shot the Kennedy, someone else would have.

Entropy: In Shannon’s theory, entropy is the measure of the uncertainty of a message. This concept, which is originated from Thermodynamics, is prevalently used in different fields and in different senses. Shannon’s entropy is the sense that is being used in this chapter.

Kullback – Leibler Divergence Measure: This information-theoretic concept is a measure of the divergence between the probability distributions of random variables. If we assume that p and q are the probability distributions of two random

variables, then the Kullback-Leibler Divergence, $D(p||q)$, is calculated with the following formula:
 $\sum p_i \cdot \log (p_i/q_i)$

Markov Chains: A Markov chain is a stochastic process with the Markov property: A process has the Markov property if the conditional probability distribution of the future states of the process depends upon only the present state and a specific (say m) number of past states. The number m determines the order of the Markov chain. For example, Markov chains of order 0 are called memoryless systems because the future states depend only on the present state. In this chapter, we use the term Markov chain as a shortcut for Markov chains of order 1.

Possible Worlds Semantics: As stated in the counterfactuals entry above, the truth condition for counterfactuals is different than the truth condition for material conditionals. Possible worlds semantics developed for specifying the truth condition for counterfactuals by Lewis and Stalnaker. Lewis defines the truth condition as following:

Random Variable: Random variables are used in probability theory. They assign numerical values to the outcomes of an experiment. Usually, they are represented by capital letters such as X, Y.

ENDNOTES

- ¹ In the title of his article, Shannon intentionally used the word ‘communication’ to avoid philosophical ambiguities of ‘information’. Despite this, the common practice in the literature is to call his theory ‘the mathematical theory of information’. This common practice is adapted in this chapter for consistency with the literature.
- ² Several people claimed that this connection between the entropy of thermodynamics

and the measure for the expected surprisal value (information) points out some deep metaphysical connections (Wiener 1961, Wheeler 1994, Chalmers 1996, Brooks & Wiley 1988].

- ³ For an analysis of and a suggested solution for this paradox, please see Floridi (2004) and Floridi’s forthcoming book, *The Philosophy of Information*.
- ⁴ Dretske’s BBS open commentary article (1983) and the special issue of *Synthese* on Dretske’s theory (1987) together give us a valuable collection of these criticisms.
- ⁵ The theorem is rephrased for the sake of simplicity.
- ⁶ For details regarding these problems, please see Demir’s dissertation (2006), which is available at <http://scholarworks.iu.edu>.
- ⁷ It is important to note that some philosophers disagree with this claim. Bar-Hillel (1952), Hintikka (1970) and Bremer (2003) are useful sources for a balanced presentation of this debate.
- ⁸ Brogaard and Salerno (2008) claim that when the contextual features of an argument are taken into account, counterfactuals satisfy transitivity. It needs to be stated that their claim is based on a misunderstanding of Lewis’ possible worlds semantics. The details of their misunderstanding will need to be explained some other time.
- ⁹ Dietrich (2007) has a concise review of such attempts.
- ¹⁰ For the sake of simplicity, I use ‘Markov chains’ for ‘Markov chains of order 1’. Markov chains can have any number of order. For details, please see the Glossary.