7-2009

Information of Interactions in Complex Systems

Klaus Krippendorff
University of Pennsylvania, kkrippendorff@asc.upenn.edu

Recommended Citation

Krippendorff, K. (2009). Information of Interactions in Complex Systems. *International Journal of General Systems, 38* (6), 669-680. https://doi.org/10.1080/03081070902993160

Information of Interactions in Complex Systems

Abstract
This paper addresses misconceptions of the multi-variate interaction-information measure Q, which several authors have reinvented since its proposal by McGill (1954), giving it a variety of names and interpretations. McGill's measure claimed to quantify the amount of information of interactions among three or more variables in complex systems. In (Krippendorff, 1980), I raised doubts about the validity of Q and its relatives. The chief problem that Q-measures fail to recognize is that complex interactions tend to involve circularities and the probability distributions characterizing such circularities cannot be obtained by products of probabilities, which underlie information theory as far as developed by Shannon (Shannon & Weaver, 1949). I argued that Q-measures are mere arithmetic artifacts, and proposed an algorithmic solution to measuring the amount of information in the interactions within complex systems, now widely accepted. The paper responds to Leydesdorff's (2009) "Interaction information: Linear and nonlinear interpretations," published in the current issue of this journal and preceding discussions of these issues on the Cybernetics Discussion Group CYBCOM and personal correspondence involving Jakulin (2009). It prefers to rely on demonstrations with numerical data over abstract interpretations of mathematical forms that can so easily entrap scholars into believing that they measure something real without considering evidence to the contrary.

Keywords
Information Theory, Complex Systems, Interaction Information, Cybernetics. Decomposition

Disciplines
Communication Technology and New Media | Organizational Communication | Other Social and Behavioral Sciences

International Journal of General Systems
Vol. 38, No. 6, August 2009, 669–680

Taylor & Francis
Taylor & Francis Group

COMMENTARY

Information of interactions in complex systems

Klaus Krippendorff*

The Annenberg School for Communication, University of Pennsylvania, 3620 Walnut Street, Philadelphia, PA 19104-6220, USA

I am responding to Leydesdorff's (2009) 'Interaction information: linear and nonlinear interpretations', published in the current issue of this journal, but want to address the larger problem of which his is just one example. His paper seeks to justify the use of multi-variate Q-measures that can be found in several literatures as I- or information-measures, but have been shown to be algebraic artifacts with questionable statistical interpretations (Krippendorff 1980, 2009). Most users of these measures, including Leydesdorff, rely on other authors who, I suggest, are seduced like I was until 1979 by their elegant algebra, without exploring what they could possibly indicate.

The quantities in question were introduced by McGill (1954) as measures of the amount of information of the interactions in complex systems. After I raised some questions regarding them in a draft of a paper by Lucio-Arias and Leydesdorff (2009) that Leydesdorff shared on the Cybernetics Discussion Group CYBCOM and personal correspondence involving Jakulin (2009), Leydesdorff (2009) acknowledged some problems with these Q-measures, but argued for theoretical interpretations of what they do, which, I maintain, do not overcome their serious shortcomings. Leydesdorff and Jakulin (2009.2.27) weigh their arguments by referring to their common use. I prefer demonstrations instead and will proceed accordingly.

We agree on the contours of Q's definition:

$$Q(A) = -H(A) = \sum_{a \in A} p_a \log_2 p_a, \tag{1}$$

$$Q_B(A) = -H_B(A) = H(A) - H(AB) = \sum_{ab \in AB} p_{ab} \log_2 \frac{p_{ab}}{p_b}, \tag{2}$$

$$\begin{aligned} Q(AB) &= Q_B(A) - Q(A) = H(A) + H(B) - H(AB) = T(A:B) \\ &= \sum_{ab \in AB} p_{ab} \log_2 \frac{p_{ab}}{p_a p_b}, \end{aligned} \tag{3}$$

*Email: kkrippendorff@asc.upenn.edu

ISSN 0308-1079 print/ISSN 1563-5104 online

DOI: 10.1080/03081070902993160
http://www.informaworld.com

$$Q(ABC) = Q_C(AB) - Q(AB) = \sum_{abc \in ABC} p_{abc} \log_2 \frac{p_{abc}}{\left[\frac{p_{ab}p_{ac}p_{bc}}{p_a p_b p_c}\right]} \tag{4}$$

$$Q(ABCD) = Q_D(ABC) - Q(ABC) = \sum_{abcd \in ABCD} p_{abcd} \log_2 \frac{p_{abcd}}{\left[\frac{p_{abc}p_{abd}p_{acd}p_{bcd}p_a p_b p_c p_d}{p_{ab}p_{ac}p_{ad}p_{bc}p_{bd}p_{cd}}\right]}, \tag{5}$$

which are recursively extendable to any number of variables. Within a set Γ of variables, Q-measures can also be defined in terms of entropies $H(X)$ as in (1):

$$Q(\Gamma) = \sum_{X \subseteq \Gamma} (-1)^{1+|\Gamma|-|X|} H(X), \tag{6}$$

where X is a subset of Γ, $|\Gamma|$ is the cardinality of Γ and $|X|$ of X. Equations (1)–(3) correspond to Shannon's (Shannon and Weaver 1949) definition of entropy H and information transmission T. The recursive extension to three or more variables, Equations (4) and (5), is McGill's (1954). I developed (6) in Ashby's 1962–1963 in seminar and used the logarithmic forms of (3)–(5) to prove their inadequacies (Krippendorff 1980). All of the following generalisations are Ashby's (1969), some of which were foreshadowed but not fully appreciated by others. Much of the literature focuses on ternary interactions that (6) generalised:

$$Q(ABC) = -H(A) - H(B) - H(C) + H(AB) + H(AC) + H(BC) - H(ABC). \tag{7}$$

Shannon defined the entropies in (1), as used in (2), (3), (6) and (7), in terms of probability distributions and information as the difference between two entropies, for example, between the entropy at a receiver $H(R)$ without reference to a sender and $H_S(R)$ by reference to the sender's choices, the uncertainty U_{before} and U_{after} a message was received, or the entropy of an observed system and the hypothetical entropy of the aggregate of its (unrelated) parts. Accordingly, the total amount of information in a system of variables is the difference between the hypothetical entropy $\sum H(X)$, that regards all variables X of a system within a set Γ of variables as independent of each other, and the observed entropy $H(\Gamma)$ in that system as a whole, i.e.:

$$T(\Gamma) = \sum_{X \in \Gamma} H(X) - H(\Gamma). \tag{8}$$

For example, within just three variables:

$$T(A{:}B{:}C) = H(A) + H(B) + H(C) - H(ABC). \tag{9}$$

This convenient information calculus led Ashby (1969) to numerous accounting equations, among which is the equality of the total amount of information in a system (8), and the sum of all of its Q-quantities (6), allowing the total complexity in a system to be decomposed into additive quantities, each notationally tied to its less complex interaction components:

$$T(:\Gamma) = \sum_{S \subseteq \Gamma} Q(S), \tag{10}$$

where S is a subset of Γ containing at least two variables. This expression is elegant and appealing for its simplicity. For three variables, (10) becomes:

$$T(A{:}B{:}C) = Q(AB) + Q(AC) + Q(BC) + Q(ABC), \tag{11}$$

the sum of all interactions in a system, here of three variables. Without always appreciating the generality of these equations for decomposing the information in a system into quantities associated with its constitutive parts, several authors found the use of Q-like measures attractive indicators of the amount of information in interaction. Jakulin (2005) reviewed numerous applications of this idea and justified his own use of Q-like measures by reference to these. But how do these measures relate to the probability distributions on which information theory is defined?

The probabilities in the numerators of (1)–(5) are observed probabilities, of course, and they yield entropies as in (1). The probabilities $p_a p_b$ in the denominator of (3) are probabilities as well, in particular, of what can be expected when variables A and B are statistically independent or yield maximum entropies. This is how far Shannon went. However, the products of probabilities in the denominators of (4) and (5), and in all higher-order Q-terms do not add to one (see Table 1 below for an example), cannot be interpreted as probabilities, do not yield entropies as in (1), making it difficult to interpret Q for three or more variables as a measure of information (Krippendorff 1980, 2009). Watnabe (1960) noticed this before I knew of his work, considered these products to have 'no profound meanings', and counselled against their use in information theory. Jakulin (2005) recognised this as well, but after personal communications, Jakulin (2009.2.26 & 27) acknowledged that while Q-measures are not perfect, they are good approximations to maximum entropies and Leydesdorff (2009) sought other explanations to justify their continued use. I consider salvaging this measure to be an exercise in futility and am extending Watnabe's judgment to all quantities that include these products. *Q-quantities may have other uses, and I will mention one below, but multivariate information measures they are not.*

But there are other reasons not to interpret Q-terms as measuring the information in multi-variate interactions. Since McGill (1954), zero quantities of Q have been interpreted as the absence of interactions (Garner and McGill 1956; Garner 1962; Matsuda 2000; Bell 2003; Jakulin 2005; Yeung 2008; Lucio-Arias and Leydesdorff 2009). This claim is demonstrably false. Consider the application of (11) to data in Figure 1, which consist of three separate frequency distributions in three dichotomous dimensions, resulting in quantities of $Q(ABC)$ of -1, 0 and $+1$ respectively:

Example C visualises a non-decomposable interaction, in fact, the strongest interaction possible in three dichotomous variables. Here, the total amount of information $T(A{:}B{:}C)$ in the system is taken up by $Q(ABC)$. All three binary interactions are absent and in this extreme case, and only then, $Q(ABC)$ quantifies the visually apparent ternary interaction in the data cube. The other extreme is found in example A. Here, $Q(ABC)$ is maximally negative. In this data cube, one may notice that any two of the three two-dimensional frequency distributions are sufficient to reconstruct the three-dimensional distribution in ABC, hence the ternary interaction is absent and any one of the three binary interactions can be dropped for being superfluous in specifying the observed frequencies. Equation (11) does not reflect this logic, however. All three binary interactions measure maximum amounts of information, revealing $Q(ABC)$ to be a left-over quantity that compensates for the failure to exclude one of the redundant binary interactions.

However, the main point of this demonstration may be seen with example B. Here, $Q(ABC) = 0$, supposedly indicating the absence of ternary interaction. Yet, the frequency distribution in this data cube obviously is far from what could be expected by chance, given the distributions observed in AB, AC and BC. The algebraic forms of (4) and (5) notwithstanding, the example provides a visual demonstration that *interpreting* $Q = 0$ *as indicating the absence of interaction is plainly mistaken when three or more variables are involved*. This statistical fact cannot be appreciated without actually observing how Q responds to different frequency distributions, as in Figure 1.

To make sense of the peculiarity of negative and positive values of Q, Leydesdorff (2009) relies on the analogy between entropies and Venn diagrams. Venn diagrams, which depict all possible intersections of several sets, are widely used in conceptualising the decomposition of entropies into all possible interactions among variables (e.g. McGill 1954; Theil 1972; Bell 2003; Yeung 2008). Accordingly, $H(A)$ is analogue to the set A, $Q(AB) = T(A{:}B)$ is analogue to the intersection $A \cap B$, $Q(ABC)$ is analogue to the intersection $A \cap B \cap C$, etc. But Leydesdorff goes beyond this analogy, locating $Q(ABC) < 0$ in the intersection $A \cap B \cap C$ inside the union $A \cup B \cup C$, but $Q(ABC) > 0$ outside the union $A \cup B \cup C$ for which no entropy is defined. Analogies can be and in this case clearly are misleading, as MacKay (2003, pp. 143–144) observed, not only because it is far from obvious how elements in sets correspond to entropies, but also because of the odd role of Q. In pursuit of this analogy and by reference to Abramson (1963) and Leydesdorff (2009) defines his measure of mutual information as:

$$I(ABC) = H(A) + H(B) + H(C) - H(AB) - H(AC) - H(BC) + H(ABC = -Q(ABC), \tag{12}$$

in effect adopting the Q-measure in (7) but with a negative sign, as if this would turn the statistically uncertain quantity $Q(ABC)$ into a proper information measure $I(ABC)$ – incidentally labelled μ^* in Lucio-Arias and Leydesdorff (2009) after Yeung (2008), and equivalent to $-A'(uvw)$ in McGill (1954) and $-I(A;B;C)$ in Jakulin (2005). Leydesdorff does not say how he would define his I for other than three variables. But because he justifies his version of Q by reference to Abramson, it might be worth noting that Abramson, Matsuda (2000), Bell (2003), Yeung (2008) and possibly others define $I(\Gamma) = Q(\Gamma)$ for even numbers of variables, and $I(\Gamma) = -Q(\Gamma)$ for odd numbers of variables. I have not found any motivation for the proposal of the odd/even reversal of the sign of Q. There is a history of this sign-reversal. Quastler (1953) started with a similar definition but shifted to McGill's Q in their joint effort to standardise the nomenclature of information theory (McGill and Quastler 1955). I am suggesting that this alternating sign-reversal does not address Q's problems, makes accounting equations for entropies unnecessarily difficult – although in all fairness, unlike Ashby (1969), Leydesdorff is not concerned with generalising accounts such as in (10) – but it demonstrates the common uncertainty about Q's meaning. Jakulin (2005, p. 37) even goes so far as to call *any* deviation from Q's zero value an 'interaction magnitude', in effect ignoring the sign and with it the differences in frequency distributions in examples A and C. This raises the question: *what could a measure possibly mean whose zero point is unrelated to the absence of interactions and for which some researchers interpret its positive values just as other researcher interpret its negative values?*

Jakulin (2005, p. 39ff) reviewed numerous studies from diverse disciplines ranging from physics, chemistry, biology, cognitive science, to economics, including by Leydesdorff and Meyer (2003), that have relied on Q in one form or another. I contend, *the*

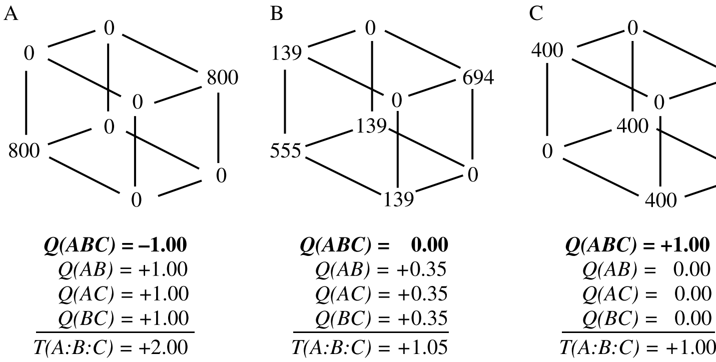


Figure 1. Accounts of three so-called interactions Q.

above demonstration calls all uses of Q-like measures for more than two variables into question.

Without examining the strange behaviour of Q vis-à-vis the probability distributions to which it responds, Leydesdorff (2009), relying on Garner and McGill's (1956) observation of parallels between their accounting information for information and the analysis of variance, concludes that 'the Q-measure is mathematically sound but an interpretation from the perspective of information theory needs to be provided'. He then cites Jakulin (2005) who called, following a suggestion by Gat (1999), positive Qs measures of synergy and negative Qs measures of redundancy. In search for a meaning for negative measures, Matsuda (2000) called them 'frustrated correlations' and frustrating they are! To find more profound meanings for positive and negative Qs, Leydesdorff (2009) then links them to the results of positive and negative feedback and to Maturana's theory of autopoiesis. Yet, circular causal feedback and self-production are dynamic and informationally closed systems (Ashby 1956) whose complexities go far beyond the ability to represent them in terms of simple multi-dimensional probability distributions, which Leydesdorff is quantifying with Q. I am suggesting that abstract theoretical interpretations cannot rescue Q from its problems. They merely obscure them. Garner and McGill observed nothing but a conceptually intriguing analogy from which no mathematical soundness can be inferred. This is not to questions the soundness of Q's definitions and algebra, but this soundness does not extend to its interpretation as an information measure. *No theoretical interpretation can change the uncertain relationship between Q and statistically evident interactions.*

What could replace the defective Qs and maintain the idea of decomposing the information in a complex system into less complex subsystems, each associated with a quantity of information they process?

To me, the answer to this question started to emerge as a result of a gestalt switch that Klir (1978) introduced by representing sets of variables in subsystems of a larger system as boxes and the variables they share as connections between them. This visualisation revealed that all systems with three or more variables can contain circular dependencies. It turned out that the probabilities in such loops can no longer be obtained by multiplying the participating components' probabilities, as undertaken in the denominators of (4) and (5). Not acknowledging these loops caused all of Q's oddities. I then used an iterative algorithm (Krippendorff 1980, 2009), further developed (Krippendorff 1986) and available at http://www.pdx.edu/sysc/research-discrete-multivariate-modeling (last accessed 7 April 2009), which goes around and around these circularities and converges on the maximum entropy probability distribution, preserving the probability distributions of the components of a system of variables. The use of this iterative algorithm also called for reconceptualising the lattice of possible decompositions – incidentally not as simple as those derived from Venn-diagrams and explored by Bell (2003) and Jakulin (2005) among others. This lattice (Krippendorff 1986, p. 40; 2009, p. 198) emerges when removing, starting from the unanalysed whole system m_0, one interaction after another, each in the context of the remaining interactions. This step-by-step removal of interactions from m_0 creates progressively simpler models $m_1, \ldots, m_i, m_{i+1}, \ldots, m_{\text{ind}}$ of m_0, m_{ind} being the model consisting of independent variables. This process defines a path through this lattice with increasing entropies $H(m_i) \leq H(m_{i+1})$. The information in any *one* interaction then became the difference between two entropies, the maximum entropy including and excluding the interaction in question:

$$I(m_i \rightarrow m_{i+1}) = H(m_{i+1}) - H(m_i) = \sum_{abc\ldots \in ABC\ldots} \omega_{abc\ldots(m_i)} \log_2 \omega_{abc\ldots(m_i)} - \sum_{abc\ldots \in ABC\ldots} \omega_{abc\ldots(m_{i+1})} \log_2 \omega_{abc\ldots(m_{i+1})} \quad (13)$$

where probabilities $\omega_{abc\ldots(m_i)}$ are the iteratively obtained maximum entropy probabilities for the model m_i of m_0, and $\omega_{abc\ldots(m_i+1)}$ for m_{i+1}. All information measures are zero or positive quantities and add to the total amount of information in a system, in effect rescuing the idea of (10) from Q's failures:

$$T(m_0) = I(m_0 \rightarrow m_{\text{ind}}) = I(m_0 \rightarrow m_1) + \cdots + I(m_i \rightarrow m_{i+1}) + \cdots + I(m_{\ldots} \rightarrow m_{\text{ind}}). \quad (14)$$

Applying (14) to two kinds of data, depicted in Figure 2, yields accounts of the information quantities involved, which can now be compared to the Q-quantities obtained by (11).

The frequency distribution in example D is the same as in example B of Figure 1. The decomposition of the total amount of information into its constitutive interactions demonstrates that the measure of $Q(ABC) = 0$ completely misses the ternary interaction, visually evident in D's data cube and measured as $I(ABC \rightarrow AB{:}AC{:}BC) = 0.25$ bits. The data in example E are derived from D by removing its ternary interaction, representing the maximum entropy frequency distribution, verifiably satisfying the original distributions in the three binary interactions AB, AC and BC. Removing that ternary interaction from D zeroes $I(ABC \rightarrow AB{:}AC{:}BC)$ in E and subtracts the corresponding 0.25 bits from E's total $I(ABC \rightarrow A{:}B{:}C) = T(A{:}B{:}C) = 0.80\,(=1.05-0.25)$ as it should, but also from the quantity $Q(ABC)$! This makes unquestionably clear that *Q cannot be interpreted as measuring a unique property of interactions*. It is affected by something else.

But what does Q actually measure? This question led me (Krippendorff 1980, p. 66) to distinguish two opposing quantities within Q, the correct amount of information I in interactions and a measure of the over determination or redundancy R in the algebraic specifications of these interactions. For the highest-order interaction in ABC but absent from $AB{:}AC{:}BC$,

$$Q(ABC) = I(ABC \rightarrow AB{:}AC{:}BC) - R(\mathrm{AB{:}AC{:}BC}). \quad (15)$$

Accordingly, Q is the difference between the interaction information I and the mistake made by what one may call a Boolean observer who insists on accounting for that interaction algebraically, as in (6) and (7), not iteratively. Being a difference, Q cannot be interpreted as a stand-alone measure of interaction, redundancy, or synergy, as Leydesdorff (2009; Lucio-Arias and Leydesdorff 2009), Jakulin (2005), and many others claim. Jakulin (2005, p. 41) may have intuited the involvement of redundancy in Q when he writes 'Negative interactions (meaning Q) imply redundancy, which may be complete or partial'. But under the mistaken assumption that $Q = 0$ measured the absence of interaction, he identified positive Qs as measures of synergy and negative Qs as measures of redundancy. Equation (15) suggests a more complicated relationship between these quantities. It does not suggest that Q 'is to be discarded as incompatible with information theory', as Leydesdorff (2009) reads me as saying. It provides to be the key to a measure of the redundancy in the algebraic specification of interactions:

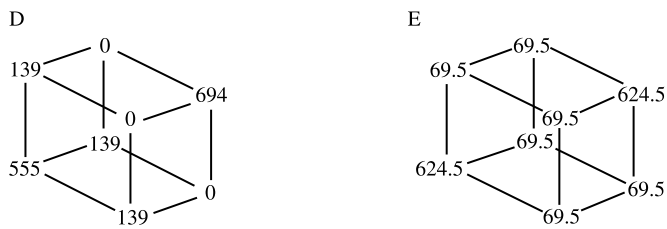
$$R(m_1) = I(m_0 \rightarrow m_1) - Q(m_0). \tag{16}$$

The notations in the arguments of I, Q, R reveal their conceptual difference. Q-quantities are defined in terms of the variables of an interaction, whereas I- and R-measures also take account of how the remaining components interact with one another. *Information quantities in* (13) *are context sensitive, Q-quantities in* (11) *are not*. The arguments in R refer to call components of a system absent the interaction that I is measuring. According to (16), in example D, $R(AB{:}AC{:}BC) = I(ABC \rightarrow AB{:}AC{:}BC) - Q(ABC) = 0.25$ bits. The first binary interaction, $I(AB{:}AC{:}BC \rightarrow AC{:}BC)$ contributes 0.10 bits of information, not $Q(AB) = 0.35$, while the second and third binary interactions contribute all of their information, 0.35 bits each, to the total.

Generalising (16) to any model m_i of m_0 and taking full advantage of (10) for each of the interactions removed from m_0, yields:

$$R(m_i) = I(m_0 \rightarrow m_i) - \sum_{K \subseteq \Gamma \text{ and } K \nsubseteq m_i} Q(K), \tag{17}$$

where K is a subset of the set Γ of a system's variables and $K \nsubseteq m_i$ prevents K from being contained in any of m_i's components. R is positive when the algebraic account exaggerates the information in m_i and negative otherwise. Leydesdorff provided an empirical example



D	E
$I(ABC \rightarrow AB{:}AC{:}BC) = 0.25 \neq Q(ABC) = 0.00$	$I(ABC \rightarrow AB{:}AC{:}BC) = 0.00 \neq Q(ABC) = -0.25$
$I(AB{:}AC{:}BC \rightarrow AC{:}BC) = 0.10 \neq Q(AB) = 0.35$	$I(AB{:}AC{:}BC \rightarrow AC{:}BC) = 0.10 \neq Q(AB) = 0.35$
$I(AC{:}BC \rightarrow A{:}BC) = 0.35 = Q(AC)$	$I(AC{:}BC \rightarrow A{:}BC) = 0.35 = Q(AC)$
$I(A{:}BC \rightarrow A{:}B{:}C) = 0.35 = Q(BC)$	$I(A{:}BC \rightarrow A{:}B{:}C) = 0.35 = Q(BC)$
$I(ABC \rightarrow A{:}B{:}C) = 1.05 = T(A{:}B{:}C)$	$I(ABC \rightarrow A{:}B{:}C) = 0.80 = T(A{:}B{:}C)$
$R(AB{:}AC{:}BC) = 0.25 - 0.00 = 0.25$	$R(AB{:}AC{:}BC) = 0.00 - (-0.25) = 0.25$
$R(AC{:}BC) = 0.35 - 0.00 - 0.35 = 0.00$	$R(AC{:}BC) = 0.10 - (-0.25) - 0.35 = 0.00$

Figure 2. Accounts of example B in Figure 1 with and without its ternary interaction.

for negative redundancies (see Jakulin 2009.4.17) Evidently, the algebraically obtained Q is not as simple and unproblematic as its proponents take it to be. Its positive and negative quantities bear a more complex relationship to $I(m_0 \rightarrow m_1)$

Note that *interactions with loops entail positive or negative redundancies, those without loops do not*. Loops can be complex, especially in systems with many variables. For example, while the model AB:AC:BC in three variables contains just one loop, A—B—C—A, the model ABC:ABD:ACD:BCD in four variables contains seven. Even a relatively simple structure, ABC:ABD:CD, contains two loops, A—C—D—A and B—C—D—B, whose redundancy, according to (17), would amount to:

$$R(ABC{:}ABD{:}CD) = I(ABCD \rightarrow ABC{:}ABD{:}CD) - Q(ABCD) - Q(ACD) - Q(BCD).$$

When removing CD from ABC:ABD:CD both loops disappear and the resulting redundancy becomes

$$R(ABC{:}ABD) = I(ABCD \rightarrow ABC{:}ABD) - Q(ABCD) - Q(ACD) - Q(BCD) - Q(CD) = 0.$$

I am suggesting that accounting for the redundancy in the specifications of interactions resolves the problems of interpreting negative Q-values. As already discussed, in example A, the distribution of frequencies in ABC can be reconstructed from any two faces of the data cube, say AC and BC. Quantitatively, $I(ABC \rightarrow AB{:}AC{:}BC) = 0$ bits indicates the absence of ternary interactions. With $Q(ABC) = -1$, redundancy measures $R(AB{:}AC{:}BC) = 1$ bit, which accounts for the redundant binary interaction in AB. Removing AB from $AB : AC : BC$ yields $R(AC{:}BC) = I(ABC \rightarrow AC{:}BC) - Q(ABC) - Q(AB) = 0$, which quantitatively accounts for the fact that AC and BC are sufficient to reconstruct ABC. In example D, the information $I(ABC \rightarrow AB{:}AC{:}BC)$ and redundancy $R(AB{:}AC{:}BC)$ are 0.25 bits each, cancelling each other in Q, explain why $Q(ABC) = 0$, and demonstrate, as already discussed, Q's failure to reveal the existence of the ternary interaction present in the data. Example C exhibits no redundancy whatsoever, $R(AB{:}AC{:}BC) = 0$, and the total absence of redundancy is the only condition under which Q can be interpreted as an information measure of interaction. Thus, *the measure of redundancy in* (17) *rescues Q from obscurity but disqualifies it as a stand-alone measure of anything* so far considered. It becomes an intermediate computational step to the measure of redundancy or of the over- or under-specification of interactions by the Boolean logic of algebiaic accounts.

The above demonstrations lead me to disagree with Leydesdorff (2009) assessment that measures of information in systems with circularities are 'theoretically incommensurable with Shannon's and Ashby's program'. Shannon's axioms (Shannon and Weaver 1949, pp. 18–20), which lead him to his conception of entropy, information, and bits as their units of measurement, apply in full to interactions that are more complex than the binary ones he explored, and the idea of Ashby's accounting equations for interactions in complex systems is preserved in (14). What is to be faulted is the assumption that one could obtain maximum entropy probabilities for systems with circular dependencies by multiplying its participating components' probabilities, or obtaining information quantities of interaction algebraically, as attempted in (4)–(6). *Q results from an inappropriate use of probability theory* on which information theory is defined.

Two points remain to be made. First, since McGill, the recursively extendable definition of Q in (3)–(5), etc. has lead to a seemingly plausible interpretation, reproduced by many users of Q-like quantities. According to McGill (1954, p. 101) $Q_C(AB) - Q(AB)$ in (4) 'is the gain (or loss) in sample information transmitted between any two of the variables, due to additional knowledge of the third variable'. If C causes the interactions in

AB, or must be present for this interaction to occur, then, arguably, $Q_C(AB) > Q(AB)$, and $Q(ABC) > 0$ would be the extent to which C controls unique combinations of values in AB over and above observing AB in isolation. By the same logic, if C weakens the interaction in AB, $Q < 0$. As Jakulin (2005, p. 41) explains, 'positive interactions (meaning Q-quantities) imply that the introduction of the new attribute (C) increased the amount of dependence (between A and B). A disappearance of a dependence is a kind of *negative interaction*: negative interactions imply that the introduction of the new attribute decreased the amount of dependence. If C does not affect the dependence between A and B, we say there is no 3-interaction'. This interpretation hides a hitch that may surface when examining what $Q_C(AB)$ actually measures.

Slicing the data cube in example C into any two planes, say along the values of variable C, $c \in \{0,1\} = C$, reveals two frequency distributions $ABc{=}0$ and $ABc{=}1$. They could not be more different from each other, clearly demonstrating the difference that C makes for the relationship between A and B, in fact causing the two correlations to flip into their opposites. $Q_{c=0}(AB)$ and $Q_{c=1}(AB)$ measure 1 bit each and average to $Q_C(AB) = 1$ bit as well. With $Q(AB) = 0$, $Q(ABC) = Q_C(AB) = 1$ bit. But note that $Q_C(AB)$ is an average amount of information. It does not respond to whether the two distributions $ABc{=}0$ and $ABc{=}1$ are same or different, that is, whether they are independent of C or change with C. In this example, the interpretation seems to work, but only because redundancy is absent, as discussed above.

To see what happens in a less perfect interaction, consider the data cube from example D. For any one variable, say again C, the frequency distributions in slice $ABc{=}0$ and slice $ABc{=}1$ are unequal as well, also demonstrating how variations in C affect AB. Even quantitatively, $Q_{c=0}(AB) = 0.65$ bits is unlike $Q_{c=1}(AB) = 0.05$ bits, but they average to $Q_C(AB) = \frac{1}{2}(0.65 + 0.05) = 0.35$ bits. Since $Q(AB) = 0.35$ bits as well, $Q(ABC) = Q_C(AB) - Q(AB) = 0$. Evidently, $Q_C(AB)$ fails to recognise C's obvious correlation with unique combinations of frequencies in AB, quite unlike what the common interpretation of (4) alleges. Why? Averages do not respond to variability. $Q_C(AB)$ wipes out the very variability on which evidence of C's effect on AB relies. The difference between the distributions in $ABc{=}0, ABc{=}1$. and AB is captured by $I(ABC \rightarrow AB{:}AC{:}BC)$ but not by $Q_C(AB)$. Thus, the interpretation of conditional Q-measures as the extent to which interactions depend on an additional variable is true only (a) when *redundancy happens to be absent*, and (b) *on the average* – whatever an average dependency means. Thus, *the common interpretation of the difference $Q_C(AB) - Q(AB)$ in* (4) *obscures the substance of its claim* (to respond to an increase or decrease in interaction due to the effect on another variable). While (1)–(3) define proper entropy and information quantities, *recursively extending Q to three or more variables*, as suggested in (4) and (5), etc., *recursively obscures the interpretability of Q beyond the sense it makes in* (3).

Finally, I wish to dispel the claim that Q is a useful approximation to the amount of information in interactions. Jakulin (2005) conducted numerous simulations comparing various interaction measures to each other and concluded, reiterated in Jakulin (2009.2.27), that Q-like measures are useful approximations of interaction information, so close as to *declare* them information measures, labelling them $I(A; B; C, \ldots)$, and thereby erasing the necessary doubt in their uncritical users' minds.

Consider just one numerical example, the observed frequencies n_{abc} in example B of Figure 1 or D of Figure 2, tabulated in Table 1.

Table 1 also lists the probabilities $p_{abc(m_0)} = n_{abc}/n$ of the original data or in model $m_0 = ABC$ and the maximum entropy probabilities $\omega_{abc(m_1)}$ in model $m_1 = AB{:}AC{:}BC$, which omits the ternary interaction potentially present in m_0. The probabilities in m_1 may

Table 1. Analysis of example B in Figure 1, treating $Q(ABC)$ as approximation to $I(ABC \rightarrow AB : AC : BC)$.

$abc \in ABC$	Observed n_{abc}	Observed $p_{abc(m_0)} = \frac{n_{abc}}{1666}$	Maximum entropy $\omega_{abc(m_1)}$	Denominator of $Q(ABC)$ $\left[\frac{p_{ab}p_{ac}p_{bc}}{p_a p_b p_c}\right]$	KSA approximation : max. entropy $\frac{1666}{1.2960}\left[\frac{p_{ab}p_{ac}p_{bc}}{p_a p_b p_c}\right] : 1666\ \omega_{abc(m_1)}$
000	139	0.0834	0.0417	0.0232	29.8235 : 69.4722
001	0	0	0.0417	0.0232	29.8235 : 69.4722
010	0	0	0.0417	0.0232	29.8235 : 69.4722
011	694	0.4166	0.3749	0.5784	743.5295 : 624.5834
100	555	0.3331	0.3749	0.5784	743.5295 : 624.5834
101	139	0.0834	0.0417	0.0232	29.8235 : 69.4722
110	139	0.0834	0.0417	0.0232	29.8235 : 69.4722
111	0	0	0.0417	0.0232	29.8235 : 69.4722
Sums	1666	1.0000	1.0000	1.2960	1666 : 1666

also be examined as frequencies, rounded for convenience, in the data cube in example E of Figure 2. The next column lists the denominators of $Q(ABC)$ from (4). If Q would be an information measure, this denominator would have to be a probability. Since it sums to 1.2960, not 1.0000, it is not, as already noted. The frequency distribution that Q implies, also called the normalised Kirkwood superposition approximation (KSA, Jakulin 2005, pp. 60–61), can now be compared to the maximum entropy distribution of frequencies it is claimed to approximate. These two sets of frequencies are tabulated side by side in the last column of Table 1. Obviously, they are far from similar. With a $\chi^2 = 181, v = 3$ degrees of freedom, the null-hypothesis that these distributions are the same has to be rejected at a level of significance $p = 0.0001$. This finding, admittedly for just one but nevertheless quite ordinary example, is extraordinarily conclusive and *recommends rejecting the hypothesis that Q-like quantities approximate interaction information measures.*

Jakulin (2009.2.27) argues that Q-like measures have proven useful in numerous applications and cites an impressive number of reinventions of Q (2005, p. 39ff). However, incidences of use do not establish validity. Validity criteria might include evidence that Q-like measures are predictive of something worth knowing or advance the understanding of a phenomenon for which other measures are lacking. To my knowledge, such demonstrations have not been provided. Mere quantifications mean nothing without additional evidence.

It is unclear to me why Jakulin, who is cognizant of algorithms for calculating maximum entropy distributions (Jakulin 2005, p. 61ff), settled on measures that behave so oddly. With the availability of faster and more powerful computers, calculating Q-like measures algebraically is no longer an important convenience over calculating $I(m_i \rightarrow m_j)$ iteratively. Jakulin (2009.2.26) granted that much without noticing two opposing measures of information and redundancy in Q, known for some time to be due to circularities in higher-order interactions. I suspect that the promoters of Q-like measures manifest the predilection of mathematically inclined researchers for elegant calculi, shying away the somewhat tedious examinations of what they indicate empirically.

Acknowledgement

I like to acknowledge valuable comments by Martin Zwick on an earlier draft of this paper and Loet Leydesdorff for data causing redundancy to be negative.

References

Abramson, N., 1963. *Information theory and coding*. New York: McGraw-Hill.

Ashby, W.R., 1956. *An introduction to cybernetics*. London: Chapman & Hall.

Ashby, W.R., 1969. Two tables of identities governing information flows within large systems. *American society for cybernetics communications*, 1, 3–8.

Bell, A.J., 2003. The co-information lattice. Available from: http://www.rni.org/bell/nara4.pdf [Accessed 20 March 2009].

Garner, W.R., 1962. *Uncertainty and structure as psychological concepts*. New York: Wiley.

Garner, W.R. and McGill, W.J., 1956. The relation between information and variance analyses. *Psychometrika*, 21, 219–228.

Gat, I., 1999. *Statistical analysis and modeling of brain cells' activities*. Thesis (PhD). Jerusalem: Hebrew University. Cited in Jakulin (2005).

Jakulin, A., 2005. *Machine learning based on attribute interactions*. Thesis (PhD). Ljubljana: University of Ljubljana. Available from: http://stat.columbia.edu/~jakulin/Int/jakulin05phd.pdf [Accessed 19 March 2009].

Jakulin, A., 2009. Personal communication involving Loet Leydesdorff and myself in the Cybernetics Discussion Group CYBCOM. Available from: https://hermes.gwu.edu/archives/-cybcom.html [Accessed 20 March 2009].

Klir, G.J., 1978. Structural modeling of indigenous systems. *Proceedings of the 22nd Annual SGSR Meeting*. Washington, DC, 13–15 February, 151–155.

Krippendorff, K., 1980. *Q:* An interpretation of the information theoretical *Q*-measures. *Fifth European Meeting for Cybernetics and Systems Research*, Vienna, April 1980. Also in R. Trapple, G. Klir and F. Pichler, eds. 1982, *Progress in cybernetics and systems research*. Vol. VIII. New York: Hemisphere, 63–67.

Krippendorff, K., 1986. *Information theory: structural models for qualitative data*. Beverly Hills, CA: Sage Publications.

Krippendorff, K., 2009. W. Ross Ashby's information theory: History, some solutions, and what we face today. *International journal of general systems*, 38, 189–212.

Leydesdorff, L., 2009. Interaction information: linear and nonlinear interpretations. *International journal of general systems*, 38, 681–685.

Leydesdorff, L. and Meyer, M., 2003. The triple helix of university-industry-government relations. *Scientometrics*, 58, 191–203.

Lucio-Arias, D. and Leydesdorff, L., 2009. The dynamics of exchanges and references among scientific texts, and the autopoiesis of discursive knowledge. *Journal of Informetrics*, 3, 261–271.

MacKay, D.J.C., 2003. *Information theory, inference, and learning algorithms*. New York: Cambridge University Press.

Matsuda, H., 2000. Physical nature of higher-order mutual information: Intrinsic correlation and frustration. *Physical review E*, 62, 3096–3102.

McGill, W.J., 1954. Multivariate information transmission. *Psychometrika*, 19, 97–116.

McGill, W.J. and Quastler, H., 1955. Standardized nomenclature: An attempt. *In*: H. Quastler, ed. *Information theory in psychology: Problems and methods*. Glencoe, IL: Free Press, 83–92.

Quastler, H., ed., 1953. *Information theory in psychology: Problems and methods*. Glencoe, IL: Free Press.

Shannon, C.E. and Weaver, W., 1949. *The mathematical theory of communication*. Urbana, IL: University of Illinois Press.

Theil, H., 1972. *Statistical decomposition analysis*. Amsterdam: North-Holland Publishing Company.

Watnabe, S., 1960. Information theoretical analysis of multivariate correlation. *IBM journal of research and development*, 4, 66–86.

Yeung, R.W., 2008. *Information theory and network coding*. New York: Springer. Available from: http://iest2.ie.cuhk.edu.hk/~whyeung/post/main2.pdf [Accessed 19 March 2009].