Applications of Information Theory to Perceptual Independence and Separability

Mikaela Akrenius (makreniu@indiana.edu)
Cognitive Science Program, 1001 E. 10th Street, Bloomington, IN 47405 USA

Abstract

Despite of strong historical connections between information theory and the study of perceptual independence and separability, few modern approaches take advantage of these connections. We revive Garner and Morton's (1969) classic Mutual Uncertainty Analysis (MUA), complement it with Partial Information Decomposition (PID, Williams & Beer, 2010), and apply both to a sample of data from contemporary studies. While existing theories can dissociate between perceptual and decisional separability and identify dependencies at the level of individual stimuli, MUA and PID can provide diagnostics for identifying other types of perceptual dependencies, decompose them into their constituents, and provide a measure for their strength.

Keywords: perceptual independence; perceptual separability; information theory; mutual uncertainty analysis; partial information decomposition; general recognition theory

Introduction

Originating in studies of selective attention (Stroop, 1935) and building on Garner's (1974) speeded classification paradigm, the study of perceptual independence and separability has become a field of its own (see Algom & Fitousi, 2016, for a review). Over the recent decades, articles and book chapters on Garner interference have come to be dominated by roughly two kinds of modeling approaches: multidimensional, signal-detection-based theories, such as general recognition theory (GRT, Ashby & Townsend, 1986), and similarity- or distance-based approaches, such as the similarity choice model (Luce, 1963; Shepard, 1957) and its further extensions (e.g. Nosofsky, 1985).

Regardless of the modeling approach used, assessments of perceptual independence are typically made based on confusion matrix data from identification experiments, and build on the assumption that the distribution of response errors is diagnostic of types of violations of perceptual independence.

As Algom and Fitousi (2016) note, despite of the strong connections that Garner's (1962) early work on perceptual independence has to information theory, and of the usefulness of information theory in quantifying types of dependencies, it has seldom been used in the field.

To help cover this gap and to investigate whether and how information theory could be used, we will (1) re-introduce Garner & Morton's (1969) classic mutual uncertainty analysis (MUA), along with information-theoretic preliminaries, (2) extend it with Partial Information Decomposition (PID, Williams & Beer, 2010), (3) apply both to identification experiment data from contemporary studies and compare the results to existing, more commonly applied diagnostics (GRT), and (4) provide tentative psychological interpretations for the terms associated with PID.

Throughout the paper, we will highlight some of the formal connections between MUA, PID, and GRT. Due to limited space, this analysis will be illustratory rather than axiomatic.

Terminology

In this paper, 'perceptual independence' will be used to refer to the existence of statistical independence between the perceptual effects of (orthogonal) stimulus components. This is in line with Garner and Morton's (1969) use of the term and the definition of perceptual independence used in GRT. 'Perceptual separability', on the other hand, will be used to refer to perceptual separability as defined by GRT.

Preliminaries: Entropy, Conditional Entropy, and Mutual Information[1]

As McGill (1954) and Garner (1962) note, mutual information is an efficient tool for assessing statistical independence between two or more random variables. Unlike uncorrelation, the lack of mutual information implies statistical independence, and mutual information can capture complex (e.g. nonlinear) dependencies between variables.

Let p_i, $i \in [1, \ldots, n]$, and p_j, $j \in [1, \ldots, m]$, denote the probability associated with each of n, m outcomes of a discrete random variable x, y, respectively. The Shannon (1948) entropy of x is

$$U(x) = -\sum_{i=1}^{n} p_i \log_2 (p_i), \tag{1}$$

the joint entropy of x and y is

$$U(x,y) = -\sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} \log_2 (p_{ij}), \tag{2}$$

and the conditional entropy of x given y is

$$U_y(x) = U(x,y) - U(y). \tag{3}$$

The mutual information[2] between x and y is

$$U(x{:}y) = U(x) + U(y) - U(x,y), \tag{4}$$

the mutual information between x and two discrete random variables y, z, or three discrete random variables y, z, w, is

$$U(x{:}y,z) = U(x) + U(y,z) - U(x,y,z) \tag{5}$$

$$U(x{:}y,z,w) = U(x) + U(y,z,w) - U(x,y,z,w) \tag{6}$$

[1] Unless otherwise noted, the definitions used in this chapter are borrowed from Garner (1962) and McGill (1954).

[2] Also referred to as partial contingent uncertainty (Garner, 1962; Garner & Morton, 1969) or transmitted information (McGill, 1954).

and the conditional mutual information between x and y given z, or given z and w, is

$$U_z(x{:}y) = U(x{:}y,z) - U(x{:}z) \tag{7}$$

$$U_{zw}(x{:}y) = U(x{:}y,\ z,\ w) - U(x{:}z,w). \tag{8}$$

Mutual information is a symmetric measure of association: it is 0 if and only if x and y are statistically independent, and it can be expressed as the Kullback-Leibler (1961) divergence of the joint distribution (x, y) from the product of their marginal distributions

$$U(x{:}y) = \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}\, log_2\left(\frac{p_{ij}}{p_i p_j}\right). \tag{9}$$

The mutual information between a target variable x and two source variables y, z can also be defined as

$$U(x{:}y,z) = U(x{:}y) + U(x{:}z) + U(xyz) \tag{10}$$

where $U(xyz)$ denotes interaction information. Interaction information is a symmetric measure

$$\begin{aligned} U(xyz) &= U_x(y{:}z) - U(y{:}z) \\ &= U_y(x{:}z) - U(x{:}z) \\ &= U_z(y{:}z) - U(y{:}z) \end{aligned} \tag{11}$$

and can be interpreted as a measure of effect size.

Mutual Uncertainty Analysis (MUA)

Garner and Morton (1969) decompose the mutual information between two stimulus components A, B (e.g. shape and color) and two response variables a, b into

$$U(a,\ b{:}A,\ B) = U(a{:}b{:}A{:}B) - U(a{:}b) \tag{12}$$

$$\begin{aligned} U(a{:}b{:}A{:}B) &= U(A{:}B) + U(a{:}A) + U(b{:}B) \\ &+ U_A(a{:}B) + U_B(b{:}A) + U_{AB}(a{:}b) \end{aligned} \tag{13}$$

where $U(A{:}B) = 0$ for orthogonally varied components, $U(a{:}A)$ and $U(b{:}B)$ measure the accuracy of responses on each component ($U(a{:}A) = U(A)$ and $U(b{:}B) = U(B)$ for maximum accuracy), and perceptual independence is violated if $U_A(a{:}B) \neq 0$, $U_B(b{:}A) \neq 0$, or $U_{AB}(a{:}b) \neq 0$. According to Garner and Morton, $U_A(a{:}B) \neq 0$ and $U_B(b{:}A) \neq 0$ reflect a crossing over from one perceptual channel to the other, whereas $U_{AB}(a{:}b) \neq 0$ measures error correlation, which can be due to perceptual or response processes. Error correlation can reflect state correlation, i.e. variation in responses across trials caused by changes in the state of the observer relative to the processing channels. $U(a{:}b)$, on the other hand, reflects response correlation. These terms are illustrated in Figure 1.

Partial Information Decomposition (PID)

Partial Information Decomposition (Williams & Beer, 2010) decomposes the interaction information between one target variable and two or more source variables into redundant and synergistic components, which, intuitively speaking, reflect the information shared by the sources for predicting the target (analogous to an AND gate), and unique combinations of the sources for predicting the target (analogous to a XOR gate). Formally, the information shared between target x and sources y, z can be broken into

$$\begin{aligned} U(x{:}y,z) &= U(x{:}y) + U(x{:}z) + U(xyz) \\ &= U(x{:}y) + U(x{:}z) + U(x{:}\{yz\}) - U(x{:}\{y\}\{z\}) \\ &= U(x{:}\{y\}) + U(x{:}\{z\}) + U(x{:}\{yz\}) + U(x{:}\{y\}\{z\}) \end{aligned} \tag{14}$$

where $U(x{:}\{y\})$ and $U(x{:}\{z\})$ denote unique information contributed by each of the sources, $U(x{:}\{y\}\{z\})$ denotes redundant information, and $U(x{:}\{yz\})$ denotes synergistic information. $U(x{:}\{y\}\{z\}) = U_{min}(x{:}\{y,z\})$, the minimum amount of information shared by y and z for predicting x, and $U(x{:}y) = U(x{:}\{y\}) + U(x{:}\{y\}\{z\})$. This partitioning is illustrated in Figure 2 and can be further extended to any number of source variables. Figure 3 illustrates the case for three source variables and one target variable.

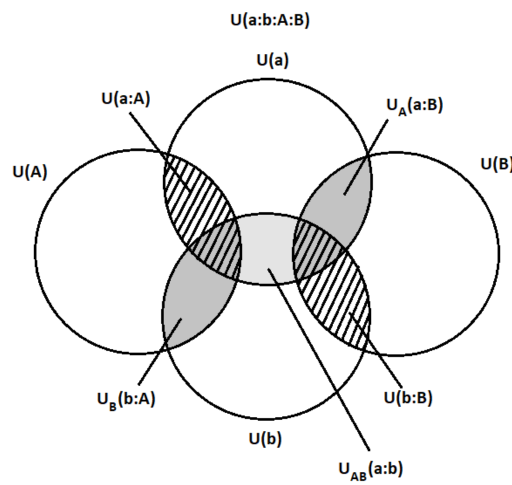

Figure 1: Illustration of the terms used by Garner and Morton (1969) in mutual uncertainty analysis.

PID for Identification Experiment Data

Identification experiment data typically involves as many response dimensions as stimulus dimensions, whereas PID has been developed to predict only one target. Due to this, PID needs to be applied separately to each response dimension. Because a majority of identification experiments consist of only two dimensions, this is relatively straightforward, and provides insight on asymmetric dependencies between the response dimensions.

Applying PID to predict response variable a yields[3]

[3] PID could also be applied to decompose the sole influence of A and B on a or b ($U(a{:}A,B)$ and $U(b{:}A,B)$), but this would yield less information and neglect interactions between the response dimensions.

$$
\begin{aligned}
&U(a{:}\,b,A,B) \\
&= U(a{:}\,\{b\}) + U(a{:}\,\{A\}) + U(a{:}\,\{B\}) + \\
&U(a{:}\,\{b\}\{A\}) + U(a{:}\,\{b\}\{B\}) + U(a{:}\,\{A\}\{B\}) + \\
&U(a{:}\,\{b\}\{A\}\{B\}) + U(a{:}\,\{bA\}) + U(a{:}\,\{bB\}) + \\
&U(a{:}\,\{AB\}) + U(a{:}\,\{bAB\}) + U(a{:}\,\{b\}\{AB\}) + \\
&U(a{:}\,\{bA\}\{B\}) + U(a{:}\,\{bB\}\{A\}) + U(a{:}\,\{bA\}\{bB\}) + \\
&U(a{:}\,\{bA\}\{AB\}) + U(a{:}\,\{bB\}\{AB\}) + U(a{:}\,\{bA\}\{bB\}\{AB\})
\end{aligned} \tag{15}
$$

which consists of all possible combinations of unique, redundant, and synergistic information contributed by each source alone or together. Applying PID to predict response variable b yields an analogous partitioning with $U(b{:}\,a,A,B)$.

Connection between MUA and PID

Using PID, the terms of MUA can be decomposed into their constituents. $U(a{:}\,A)$ (or, analogously, $U(b{:}\,B)$) can be decomposed into

$$
\begin{aligned}
U(a{:}\,A) = U(a{:}\,\{A\}) + U(a{:}\,\{b\}\{A\}) + U(a{:}\,\{A\}\{B\}) + \\
U(a{:}\,\{b\}\{A\}\{B\}) + U(a{:}\,\{A\}\{bB\})
\end{aligned} \tag{16}
$$

where $U(a{:}\,\{A\})$ is indicative of unique information from A, and the remaining terms reflect the redundant information shared by A and different combinations of b and B.

Psychologically, the unique information contributed by A can be interpreted as the direct and unique perceptual influence of A on a, i.e. the part of A that is accurately reflected in a responses, not influenced by B, and not shared with b. $U(a{:}\,\{b\}\{A\})$ and $U(a{:}\,\{b\}\{A\}\{B\})$ reflect correlation between a and b that is informed by A or A and B. $U(a{:}\,\{A\}\{B\})$ measures the redundant information in A and B that is reflected in response a, which should be 0 for orthogonal stimulus dimensions. Finally, $U(a{:}\,\{A\}\{bB\})$ reflects trials in which response a correlates with A and b is informed by an interaction of A and B.

Decomposing $U_A(a{:}\,B)$ (or, analogously, $U_B(b{:}\,A)$) yields

$$
\begin{aligned}
U_A(a{:}\,B) &= U(a{:}\,\{B\}) + U(a{:}\,\{AB\}) \\
&+ U(a{:}\,\{b\}\{B\}) + U(a{:}\,\{b\}\{AB\}) + U(a{:}\,\{bA\}\{B\}) \\
&+ U(a{:}\,\{bB\}\{AB\}) + U(a{:}\,\{bA\}\{AB\}) \\
&+ U(a{:}\,\{bA\}\{bB\}\{AB\})
\end{aligned} \tag{17}
$$

where $U(a{:}\,\{B\})$ is indicative of pure crossing over across perceptual channels, $U(a{:}\,\{AB\})$ reflects the synergistic influence of A and B on a, $U(a{:}\,\{b\}\{B\})$ reflects response correlation informed by B, $U(a{:}\,\{b\}\{AB\})$ reflects response correlation informed by synergistic combinations of A and B, and the remaining terms reflect types of state correlation.

The decomposition of error correlation, $U_{AB}(a{:}\,b)$, yields

$$
\begin{aligned}
U_{AB}(a{:}\,b) = U(a{:}\,\{b\}) + U(a{:}\,\{bA\}) + U(a{:}\,\{bB\}) \\
+ U(a{:}\,\{bAB\}) + U(a{:}\,\{bA\}\{bB\})
\end{aligned} \tag{18}
$$

where $U(a{:}\,\{b\})$ reflects unique information shared by a and b (due to pure response correlation, e.g. bias), and the remaining terms reflect different types of state correlation: $U(a{:}\,\{bA\})$ reflects cases in which the perception of A is enhanced (or impaired) by a certain state relative to B, $U(a{:}\,\{bB\})$ reflects cases in which B leaks into the perception of a when the observer is in a certain state relative to B, $U(a{:}\,\{bA\}\{bB\})$ reflects the redundant information shared by these cases, and $U(a{:}\,\{bAB\})$ reflects cases in which synergistic information from A and B interacts with the state of the observer, producing error correlation. Hence, under PID, $U(a{:}\,\{bAB\})$ is the term that corresponds most closely to Garner and Morton's interpretation of $U_{AB}(a{:}\,b)$.

Finally, response correlation, $U(a{:}\,b)$, can be decomposed into

$$
\begin{aligned}
U(a{:}\,b) = U(a{:}\,\{b\}) + U(a{:}\,\{b\}\{A\}) + U(a{:}\,\{b\}\{B\}) \\
+ U(a{:}\,\{b\}\{AB\}) + U(a{:}\,\{b\}\{A\}\{B\})
\end{aligned} \tag{19}
$$

where $U(a{:}\,\{b\})$ is shared with $U_{AB}(a{:}\,b)$, $U(a{:}\,\{b\}\{A\})$ and $U(a{:}\,\{b\}\{A\}\{B\})$ are shared with $U(a{:}\,A)$, and $U(a{:}\,\{b\}\{B\})$ and $U(a{:}\,\{b\}\{AB\})$ are shared with $U_A(a{:}\,B)$.

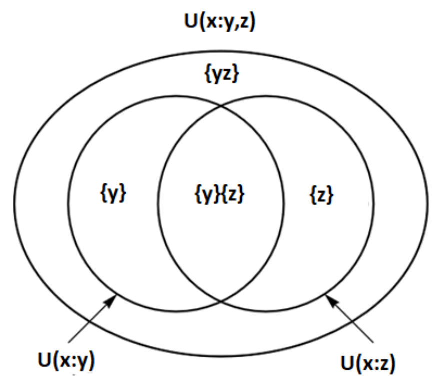

Figure 2: Partial Information Decomposition for one target variable x and two source variables y, z. Based on Figure 1 in Williams & Beer (2010).

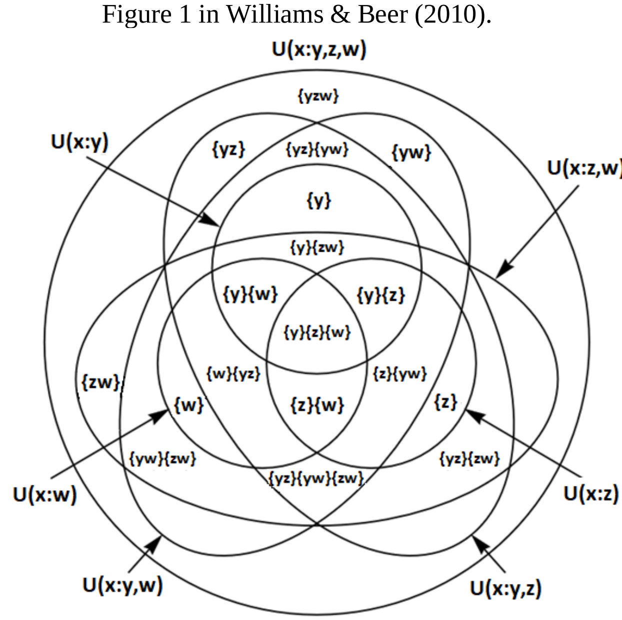

Figure 3: Partial Information Decomposition for one target variable x and three source variables y, z, w. Based on Figure S2 in Williams & Beer (2010).

Example Application: GRT

General Recognition Theory (GRT, Ashby & Townsend, 1986) is a multidimensional extension of signal detection theory, which presumes that perceptions of the stimulus components are influenced by normally distributed noise, and that response probabilities are determined by the location of choice boundaries in the perceptual space. GRT dissociates between perceptual separability, decisional separability, and perceptual independence, summarized in Figure 4, and uses a combination of probabilistic diagnostic tests to assess whether they have been violated.

According to GRT, perceptual separability holds if the perceptual distribution of a stimulus component (e.g. A_1) is uninfluenced by variation in the value of the other component (B_1 or B_2), decisional separability holds if the probability of responding e.g. a_1 given the perceptual distributions of A_1 and A_2 is uninfluenced by the value of B (i.e. the decision boundary is parallel to the B-axis), perceptual independence holds at the stimulus level if the perceptual distribution of e.g. A_1 in A_1B_1 is uncorrelated with the perceptual distribution of B_1, and perceptual independence holds at the marginal level if it is not violated for any stimulus.

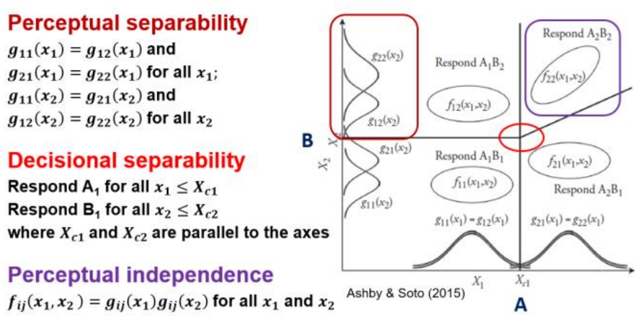

Figure 4: Postulates of GRT. Figure adapted from Ashby & Soto (2015).

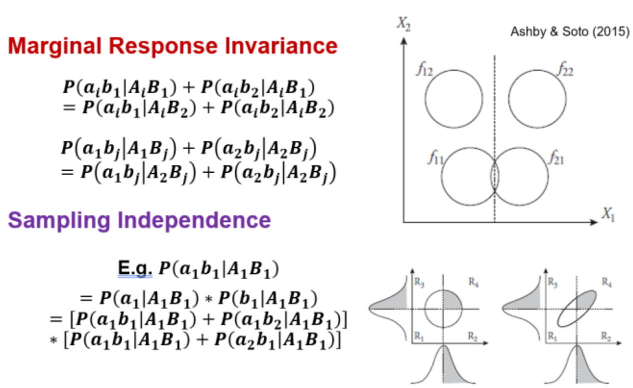

Figure 5: Some of the diagnostic tests of GRT. Figures adapted from Ashby & Soto (2015).

The diagnostic tests of GRT that are relevant for our purposes are marginal response invariance and sampling independence, summarized in Figure 5. In GRT, marginal response invariance is used as an indicator of a violation of perceptual or decisional separability, whereas sampling independence is used to assess violations of perceptual independence. GRT also employs various other statistical and signal-detection-based measures, which for the sake of space and relevance will not be reviewed here. An interested reader can consult Ashby & Soto (2015) for an illustrative review.

Formal Connections: GRT and MUA

Earlier on, it has been shown that if perceptual and decisional separability hold, marginal response invariance holds, and $U_A(a{:}B) = U_B(b{:}A) = 0$ (Theorem 6, Ashby & Townsend, 1986). Conversely, if $U_A(a{:}B) \neq 0$ or $U_B(b{:}A) \neq 0$, marginal response invariance is violated, and either perceptual or decisional separability is violated.

Analogously, it can be shown[4] that if sampling independence holds, $U_{AB}(a{:}b) = 0$ (and, conversely, if $U_{AB}(a{:}b) \neq 0$, sampling independence is violated).

Given that Ashby & Townsend (1986, Theorem 1) show that if (and only if) decisional separability holds, a violation of sampling independence implies a violation of perceptual independence (and vice versa), this means that if decisional separability holds, $U_{AB}(a{:}b) \neq 0$ indicates a violation of perceptual independence at the stimulus level.

Taken together, when $U_A(a{:}B) \neq 0$ or $U_B(b{:}A) \neq 0$, perceptual or decisional separability is violated, and when $U_{AB}(a{:}b) \neq 0$ and decisional separability holds, perceptual independence (in the GRT sense) is violated. Hence, MUA cannot dissociate between violations of perceptual and decisional separability or prove that perceptual independence (in the GRT sense) has been violated; however, MUA can provide diagnostics for other types of violations of perceptual independence.

Empirical Results: GRT, MUA, and PID

Figure 6 shows examples of GRT, MUA, and PID applied to three kinds of identification experiment data: data from one participant in a line perception study (Townsend, Hu, & Ashby, 1981), simulated data (Ashby & Soto, 2015), and data from four participants in a facial feature perception study (Thomas, 2001b). In each data set, A and B are varied in two levels, yielding a 4x4 confusion matrix of every possible combination of A and B and their respective responses.

As predicted by the formal results, when GRT indicates a violation of perceptual or decisional separability, the respective term in MUA ($U_A(a{:}B)$ for A and $U_B(b{:}A)$ for B) deviates significantly from zero in every case except for $U_A(a{:}B)$ in the Townsend, Hu, and Ashby (1981) data, which is only significant at the $p < 0.10$ or $p < 0.25$ level (depending on the correction method used). Similarly, when GRT's

[4] McGill (1954) shows that if $n_{ijm} = (n_{ij}n_{im})/n_i$ for the observed frequencies n_i, n_j, and n_m of the random variables u, v, and y, respectively, then $U_u(v{:}y) = 0$. Similarly, if $n_{ijkm} = (n_{ijk}n_{ikm})/n_{ik}$, where n_k refers to the observed frequencies of another random variable x, then $U_{ux}(v{:}y) = 0$.

perceptual independence is violated, $U_{AB}(a{:}\,b)$ deviates significantly from zero in all data sets, except for observer 4 in Thomas (2001b) where a very high accuracy for A pulls almost all interactional terms to zero. In all cases, the magnitudes of the MUA terms reflect the severity and/or number (for perceptual independence) of GRT violations.

As for the results of PID, it appears that, throughout all data sets, certain components in the decompositions of MUA terms are present very often, whereas others are seldom or never present. For example, $U(a{:}\,A)$ (and conversely $U(b{:}\,B)$) is decomposed into a nonzero $U(a{:}\,\{A\})$ (or $U(b{:}\,\{B\})$) and $U(a{:}\,\{b\}\{A\})$ ($U(b{:}\,\{a\}\{B\})$) in nearly every data set, indicating a unique contribution from A to a (B to b) and a response correlation between a and b informed by A (B). The only data sets lacking $U(a{:}\,\{A\})$ (or $U(b{:}\,\{B\})$) are Ashby and Soto (2015), where the failure of perceptual separability in A drives $U(b{:}\,\{B\})$ into $U(b{:}\,\{a\}\{B\})$ and $U(b{:}\,\{B\}\{aA\})$, i.e. all information from B is also shared with a, and observer 1 in Thomas (2001b), where $U(a{:}\,\{A\})$ is zero due to a very low accuracy in A.

Some of the data sets include a nonzero $U(a{:}\,\{b\}\{A\}\{B\})$ (or $U(a{:}\,\{b\}\{A\}\{B\})$), reflecting a crossing over in perceptual channels together with correlated responses. When both of these terms occur, the results of GRT are symmetric, whereas when only one of them occurs also GRT reflects an asymmetricity in processing. For instance, in the Ashby and Soto (2015) data, only $U(a{:}\,\{b\}\{A\}\{B\})$ is nonzero and perceptual separability is violated for A (reflecting a difference in processing across levels of B), whereas for observer 1 in Thomas (2001b) only $U(a{:}\,\{b\}\{A\}\{B\})$ is again nonzero and perceptual independence fails at only one level of B. In addition, either $U(a{:}\,\{A\}\{bB\})$ or $U(b{:}\,\{B\}\{aA\})$ is nonzero in four of the six data sets, which would appear to reflect a violation of perceptual separability in the Ashby and Soto (2015) data set but is harder to explain in the Thomas (2001b, observers 2, 3, and 4) data sets. Finally, as expected with orthogonal stimulus components, $U(a{:}\,\{A\}\{B\})$ (or $U(b{:}\,\{A\}\{B\})$) is always zero.

As mentioned earlier, $U_A(a{:}\,B)$ and $U_B(b{:}\,A)$ deviate significantly from zero only in data sets in which decisional or perceptual separability is violated. In the first case (Townsend, Hu, & Ashby, 1981), $U_A(a{:}\,B)$ and $U_B(b{:}\,A)$ are decomposed into $U(a{:}\,\{AB\})$ and $U(a{:}\,\{bA\}\{AB\})$, and $U(b{:}\,\{AB\})$ and $U(b{:}\,\{aB\}\{AB\})$), whereas in the second case (Ashby & Soto, 2015), $U_A(a{:}\,B)$ consists of $U(a{:}\,\{bA\}\{AB\})$ alone. This suggests that perceptual and decisional separability could have different signatures in PID; however, the sample of data sets used here is too small to draw further conclusions on this.

In all data sets, $U_{AB}(a{:}\,b)$ is decomposed into nonzero $U(a{:}\,\{bAB\})$ and $U(b{:}\,\{aAB\})$, four of the six data sets also have nonzero $U(a{:}\,\{bA\})$ and $U(b{:}\,\{aB\})$, and one of the data sets (Ashby & Soto, 2015) has a nonzero $U(b{:}\,\{aA\}\{aB\})$. This matches with Garner and Morton's error correlation $U_{AB}(a{:}\,b)$ being primarily reflected in $U(a{:}\,\{bAB\})$ and $U(b{:}\,\{aAB\})$. The additional terms found reflect the enhanced (or impaired) perception of one dimension depending on the state of the observer relative to the value on the other, which would (together with error correlation) appear to be reflected in GRT as a stimulus-level perceptual dependency. In one of the data sets (Thomas 2001b, observer 1), the partitioning of $U_{AB}(a{:}\,b)$ is asymmetric, with nonzero $U(a{:}\,\{bB\})$ and $U(b{:}\,\{aB\})$, possibly reflecting the fact that perceptual independence is only violated at one level of B.

Finally, as for $U(a{:}\,b)$, across all data sets only the terms shared with $U(a{:}\,A)$ or $U(b{:}\,B)$ are nonzero, indicating that response correlation always reflects information that is correct in one dimension (i.e. is never based on a relation between a and b alone, or informed by purely synergistic information from A and B).

Conclusions

To summarize, the purpose of this paper was to reintroduce MUA, to complement it with PID, to compare the results gained with MUA and PID to the results of GRT, and to provide tentative interpretations for the terms of PID. It was briefly noted that certain GRT diagnostics have MUA equivalents, and that these equivalents can be further decomposed using PID, which was illustrated in a small sample of simulated and experimental data.

The results concerning MUA and GRT are mostly in line with earlier work by Fitousi (2013), who analyzed correlations between GRT parameters and MUA terms in a simulated data set, and reanalyzed three face perception data sets (Thomas, 2001a, 2001b, and Richler et al., 2008). The novel contribution of this paper, along with formal connections between sampling independence and $U_{AB}(a{:}\,b)$, is the extension of MUA with PID and its potential psychological implications.

Suggestions for Future Work

Analogously to this paper, the results of PID could be compared to other existing approaches and extended to data sets with non-orthogonal stimulus dimensions, or to stimuli that are known to be perceptually integral. The statistical foundations underlying connections between GRT and PID would also merit further elaboration, and, like GRT, PID could be used to analyze stimulus-level information.

Methodological Notes

The MUA terms reported in this paper were computed from identification experiment data using (1) – (13) implemented in a Python program, and the PID terms presented were computed using Timme et al.'s (2014) MATLAB package. Statistical significance tests for MUA terms were executed using a chi squared approximation method described in Attneave (1959) and McGill (1954), and a correction method described in Miller and Madow (1954). The results of GRT were borrowed from the respective papers.

Acknowledgments

I would like to thank two anonymous reviewers for their comments on an earlier version of this paper.

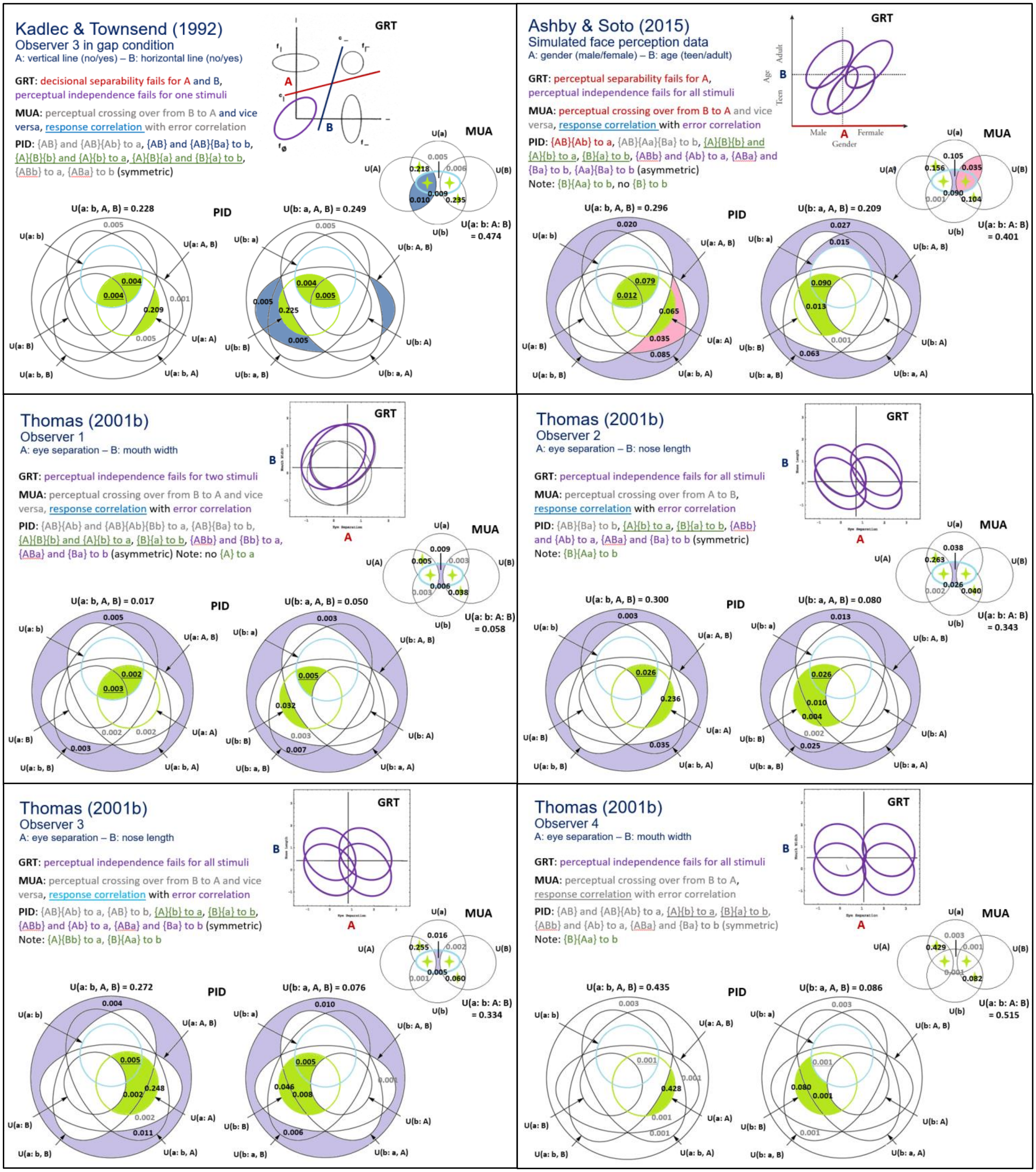

Figure 6: Examples of data sets analyzed using GRT, MUA, and PID. The figures illustrating GRT are adapted from Kadlec and Townsend (1992), Ashby & Soto (2015), and Thomas (2001b), respectively. Statistically significant MUA terms ($p < 0.05$) and their PID constituents are in black, whereas nonsignificant terms and their PID constituents are in grey. Corresponding MUA and PID terms can differ slightly due to rounding.

References

Algom, D., & Fitousi, D. (2016). Half a century of research on Garner interference and the separability-integrality distinction. *Psychological Bulletin*, *142*, 1352–1383.

Ashby, F. G., & Soto, F. (2015). Multidimensional signal detection theory. In J. R. Busemeyer, Z. Wang, J. T. Townsend, & A. Eidels (Eds.), *Oxford handbook of computational and mathematical psychology*. New York, NY: Oxford University Press.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, *93*, 154–179.

Attneave, F. (1959). *Applications of information theory to psychology: A summary of basic concepts, methods, and results.* New York, NY: Holt, Rinehart and Winston.

Fitousi, D. (2013). Mutual information, perceptual independence, and holistic face perception. *Attention, Perception, & Psychophysics*, *75*, 983–1000.

Garner, W. R. (1962). *Uncertainty and structure as psychological concepts*. New York, NY: John Wiley & Sons.

Garner, W. R. (1974). *The processing of information and structure*. Potomac, MD: Erlbaum Associates.

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

Garner, W. R., & Morton, J. (1969). Perceptual independence: Definitions, models, and experimental paradigms. *Psychological Bulletin*, 72, 233–259.

Kadlec, H., & Townsend, J. T. (1992). Implications of marginal and conditional detection parameters for the separabilities and independence of perceptual dimensions. *Journal of Mathematical Psychology*, *36*, 325–374.

Kullback, S., & Leibler, R.A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, *22*, 79–86.

Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*. New York, NY: Wiley.

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

Miller, G. A., & Madow, W. G. On the maximum likelihood estimate of the Shannon-Wiener measure of information. Air Force Cambridge Research Center: *Technical Report*, 54–75, August 1954.

Richler, J. J., Gauthier, I., Wenger, M. J., & Palmeri, T. J. (2008). Holistic processing of faces: Perceptual and decisional components. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 328–342.

Nosofsky, R. (1985). Overall similarity and the identification of separable-dimension stimuli: A choice model analysis. *Perception & Psychophysics*, *38*, 415–432.

Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, *22*, 325–345.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643–662.

Timme, N., Alford, W., Flecker, B., & Beggs, J. M. (2014). Synergy, redundancy, and multivariate information measures: An experimentalist's perspective. *Journal of Computational Neuroscience*, *36*, 119–140.

Townsend, J. T., Hu, G. G., & Ashby, F. G. (1981). Perceptual sampling of orthogonal straight line features. *Psychological Research*, *43*, 259–275.

Thomas, R. D. (2001a). Characterizing perceptual interactions in face identification using multidimensional signal detection theory. In M. J. Wenger & J. T. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges*. Mahwah: Erlbaum.

Thomas, R. D. (2001b). Perceptual interactions of facial dimensions in speeded classification and identification. *Perception & Psychophysics*, *63*, 625–650.

Williams, P. L., & Beer, R. D. (2010). Nonnegative decomposition of multivariate information. arXiv preprint, April 14th 2010.