Performance of Probability Transformations Using Simulated Human Opinions

Donald J. Bucci*, Sayandeep Acharya*, Timothy J. Pleskac†, and Moshe Kam*

*Department of Electrical and Computer Engineering, Drexel University, Philadelphia, Pennsylvania, 19104, USA
†Psychology Department, Michigan State University, East Lansing, Michigan, 48824, USA

***Abstract*—Probability transformations provide a method of relating Dempster-Shafer sources of evidence to subjective probability assignments. These transforms are constructed to facilitate decision making over a set of mutually exclusive hypotheses. The probability information content (PIC) metric has been recently proposed for characterizing the performance of different probability transforms. To investigate the applicability of the PIC metric, we compare five probability transformations (i.e., $BetP$, $PrPl$, $PrNPl$, $PrHyb$, and $DSmP$) using a simulator of human responses from cognitive psychology known as two-stage dynamic signal detection. Responses were simulated over two tasks: a line length discrimination task and a city population size discrimination task. Human decision-makers were modeled for these two tasks by Pleskac and Busemeyer (2010). Subject decisions and confidence assessments were simulated and combined for both tasks using Yager's rule and mapped into subjective probabilities using the five probability transforms. Receiver operating characteristic (ROC) curves, normalized areas under the ROC curves (AUCs), along with average PIC values were obtained for each probability transform. Our results indicate that higher PIC values do not necessarily equate to higher discriminability (i.e., higher normalized AUCs) between probability transforms. In fact, all five probability transforms exhibited nearly the same normalized AUC values. At lower, fixed false alarm rates, the $BetP$, $PrPl$, $PrNPl$, and $PrHyb$ transforms yielded higher detection rates over the $DSmP$ transform. For higher, fixed false alarm rates, the $DSmP$ transform yielded higher detection rates over the other four transforms. These trends were observed over both tasks, which suggests that the PIC may not be sufficient for evaluating the performance of probability transforms.**

***Index Terms*—Data fusion, Dempster-Shafer theory, Belief fusion, Human Simulation, Probability transformations**

I. Introduction

The Dempster-Shafer theory of beliefs is a popular tool in the information fusion community (e.g., [1]–[6]). As opposed to the use of subjective probabilities (i.e., Bayesian epistemology [7]), the Dempster-Shafer approach employs a normalized measure of evidence (i.e., belief mass assignment) on a powerset of alternatives. The result is a method of specifying imprecise evidence that results in *classes* of subjective probabilities (i.e., belief and plausibility intervals [8], [9]). To facilitate decision making, *probability transformations* are used to generate a subjective probability supported by a given belief mass assignment. There exists several Dempster-Shafer theory based fusion rules [10], as well as a large number of probability transformations (e.g., [11]–[13]). These transformations are usually evaluated through the use of hypothetical examples and by measuring the amount of entropy present in the resulting transformed probabilities for a given set of evidence (i.e., the probability information content, PIC, as in [13], [14]).

In this study, we simulate the error rates of a fusion system for a selection of probability transformations using models of human responses (i.e., decision-making, confidence assessment, and response time) from cognitive psychology. The human response model employed is the *two-stage dynamic signal detection* (2DSD) from [15]. We have used this model previously to simulate the performance of fusion combination rules over binary decision tasks in [16], [17]. In the current study, we use the *line length discrimination task* and the *city population size discrimination task* that have been previously modeled in [15]. For both tasks, human decison-makers were positioned at a computer monitor. For the line length discrimination task, subjects were presented with a pair of lines and asked to determine which of the two was longer. For the city population size discrimination task, subjects were presented with two United States cities and asked to determine which of the two had a higher population. For both tasks, subjects were asked to provide their confidences in their declarations on a subjective probability scale.

The remainder of the paper is organized as follows. Section II overviews 2DSD, the line length discrimination task, and the city population size discrimination task. Section III describes the probability transformations investigated here, and the relevant Dempster-Shafer terminology. Human responses are combined using Yager's rule [18], after which the combined results are transformed into subjective probabilities using the Pignistic (i.e., $BetP$) [19], $PrPl$, $PrnPl$, $PrHyb$ [11], and $DSmP$ [13] probability transformations. The framework of our simulation is described in Section IV. In Section V, receiver operating characteristic (ROC) curves and normalized areas under the ROC curves are estimated after applying the five probability transformations investigated here, along with the corresponding probability information content (PIC) values.

II. Human Response Simulation

The human response simulation methodology employed here involves the line length discrimination task and the city population size discrimination task given in [15]. We provide a brief overview of the 2DSD model of human responses and

the two tasks ([15] has more information on the parameter estimation and validation of the human subjects).

A. Two-Stage Dynamic Signal Detection [15]

Let $\mathcal{A} = \{A, \overline{A}\}$ represent two alternatives on a binary decision task. The 2DSD human response model simulates internal evidence accumulation for one alternative over the other, $L(t)$, using the stochastic linear difference equation,

$$\Delta L(t) = \delta \Delta t + \sqrt{\Delta t}\, \epsilon(t + \Delta t), \quad L(0) = L_0, \qquad (1)$$

where δ is the drift rate and $\epsilon(t)$ is a simulated white noise process with zero mean and variance σ^2. The parameter σ is known as the drift coefficient. The drift rate δ is positive if A is true and negative if $\overline{A}$ is true. This type of stochastic process, known as *drift diffusion*, is a common model of human decision making and response time used in cognitive psychology. To make a choice, $L(t)$ is accumulated using $\Delta L(t)$ until a threshold, either $\theta_A, -\theta_{\overline{A}}$, is crossed (where $L_0 \in [-\theta_{\overline{A}}, \theta_A]$). The decision a is then given as

$$a = \begin{cases} A & L(t) > \theta_A, \\ \overline{A} & L(t) < -\theta_{\overline{A}}, \\ \text{wait} & \text{otherwise} \end{cases}. \qquad (2)$$

Confidence assessment is achieved by waiting an additional interjudgment time, τ, and binning the final value of $L(t)$. Let $P^{(a)} = [p_1^{(a)} \cdots p_{K_a}^{(a)}]$ denote the K_a possible confidence values when choosing $a \in \mathcal{A}$ at time t_d. The chosen confidence level $p \in P^{(a)}$ for deciding a after waiting $t_c = t_d + \tau$ is given as

$$p = p_i^{(a)} \quad \text{when} \quad L(t_c) \in [c_{i-1}^{(a)}, c_i^{(a)}], \qquad (3)$$

where $c_0^{(a)} = -\infty$ and $c_{K_a}^{(a)} = \infty$ for each $a \in \mathcal{A}$. The remaining confidence bin parameters $C^{(a)} = [c_1^{(a)} \cdots c_{K_a-1}^{(a)}]$ are chosen such that $c_{i-1} < c_i$ for each $i \in \{1, \ldots, K_a - 1\}$.

The drift rate δ and initial condition L_0 can be chosen randomly at the beginning of a given simulation to allow for decision variability between trials. This randomization of δ and L_0 is performed in [15] by choosing δ from a normal distribution with mean ν and variance η^2, and choosing L_0 from a uniform distribution in the range $[-0.5s_z, 0.5s_z]$. The values ν and η are the drift rate mean and standard deviation, and s_z is the size of the interval that L_0 is chosen from. To simplify the implementation and parameter estimation of 2DSD, the authors of [15] suggest the following:

- Set $\theta_{\overline{A}} = \theta_A = \theta$.
- Standardize possible confidence assessment values (e.g., $P^{(A)} = P^{(\overline{A})} = [0.50, 0.60, \cdots, 1.00]$).
- Fix the confidence interval bins between each alternative (i.e., $C^{(A)} = C^{(\overline{A})} = C = \{c_1, c_2, \ldots, c_5\}$).
- Fix $\sigma = 0.1$.

The 2DSD parameter set for a single subject, $\mathcal{S}$ becomes

$$\mathcal{S} = \{\nu, \eta, s_z, \theta, \tau, c_1, c_2, c_3, c_4, c_5\}. \qquad (4)$$

The ten parameters defined by $\mathcal{S}$ can be determined from a subject's decision, confidence, and response time statistics using quantile maximum probability estimation [20].

TABLE I
RELATIONSHIP BETWEEN THE MEAN DRIFT RATES ν OF [15] WITH LINE LENGTH AND CITY POPULATION RANK DIFFERENCES.

Mean Drift Rate	Line length difference	Population rank difference
ν_1	0.27 mm	1 - 9
ν_2	0.59 mm	10 - 18
ν_3	1.23 mm	19 - 29
ν_4	1.87 mm	30 - 43
ν_5	2.51 mm	44 - 59
ν_6	3.15 mm	60 - 99

B. Overview of Tasks

We use the line length discrimination task and the city population size discrimination task, modeled in [15], as case studies. For the line length discrimination task of [15], six individuals were asked to compare a pair of horizontal lines with different lengths. The two lines were separated by a 20 millimeter long line. Each pair of lines consisted of a 32.00 millimeter long line and either a 32.27, 32.59, 33.23, 33.87, 34.51, or 35.15 millimeter long line. For the city population size discrimination task, six individuals were asked to compare pairs of the 100 most populated United States cities. Their answers were graded based on city population rank estimates taken from the 2006 U.S. census [21].

For both tasks, subjects were instructed to first make a declaration towards which of the two stimuli is larger (i.e. the longer line or the more populated city). Immediately thereafter the subjects were asked to assess their own confidence in that declaration on the probability scale $\{0.50, 0.60, \ldots, 1.00\}$. Subject mean drift rates ν varied based on task difficulty, as shown in Table I. The difficulty of the line length discrimination task decreases as the actual length difference between the two lines increases. The difficulty of the city population size discrimination task decreases as the difference between the population ranking of the two cities increases. Separate decision thresholds θ were determined for two cases of each task. These where when (1) subjects were asked to focus on *fast* responses; and when (2) subjects were asked to focus on *accurate* responses. Here we have used 2DSD parameter sets for subjects focusing on accurate responses. The values are available in [15, Tables 3 and 6].

III. Probability Transforms and Decision Making

A. Preliminaries

The following necessary background on Dempster-Shafer theory is taken from [8] and [10]. Consider the set of mutually exclusive alternatives Ω and its powerset 2^Ω. A Dempster-Shafer approach to information fusion assesses evidence on the powerset of alternatives through the use of *belief mass assignments* (BMAs), $m(X)$, defined for all $X \subseteq \Omega$. BMAs are normalized quantities of evidence, such that $\sum_{Z \subseteq \Omega} m(Z) = 1$. Two additional functions known as *Belief*, $\text{Bel}(X)$, and *Plausibility*, $\text{Pl}(X)$, are defined. They respectively represent a minimum and maximum amount of evidence assigned by

the BMA $m(X)$ on some $X \subseteq \Omega$. That is,

$$\text{Bel}(X) = \sum_{\substack{Z \subseteq \Omega \\ Z \subseteq X}} m(Z), \tag{5}$$

and

$$\text{Pl}(X) = 1 - \text{Bel}(\overline{X}) = \sum_{\substack{Z \subseteq \Omega \\ Z \cap X \neq \emptyset}} m(Z), \tag{6}$$

where $\emptyset$ is the empty set. When $m(\emptyset) = 0$, BMAs become normalized measures of evidence on the powerset and the belief and plausibility measures can be interpreted as bounds on potential subjective probabilities which are supported by a given BMA. For the remainder of the study, we will assume that $m(\emptyset) = 0$. The differences between the belief and plausibility bounds represents the amount of imprecise evidence given by the BMA (i.e., the mass assigned to the non-singleton elements of the powerset). If no belief masses are assigned to the non-singleton elements of the powerset, the BMA is equivalent to a subjective probability assignment.

There exists several fusion combination rules in the literature that operate using BMAs, each with its own benefits and drawbacks [10]. In the current study, we make use of *Yager's rule* for combining belief mass assignments [18]. Given two BMAs, m_1 and m_2, over the same powerset of alternatives, Yager's rule is given as

$$m_{1,2}(X) = \begin{cases} \sum_{\substack{Z_1, Z_2 \subseteq \Omega \\ Z_1 \cap Z_2 = X}} m_1(Z_1) m_2(Z_2) & X \neq \Omega \\ \sum_{\substack{Z_1, Z_2 \subseteq \Omega \\ Z_1 \cap Z_2 = X}} m_1(Z_1) m_2(Z_2) + \mathcal{K} & X = \Omega \end{cases}, \tag{7}$$

where $\mathcal{K}$ is defined as the degree of conflict between m_1 and m_2 such that

$$\mathcal{K} = \sum_{\substack{Z_1, Z_2 \subseteq \Omega \\ Z_1 \cap Z_2 = \emptyset}} m_1(Z_1) m_2(Z_2). \tag{8}$$

Yager's rule is commutative, but in general not associative [10]. To make a decision, the BMAs produced by Yager's rule must be mapped into a subjective probability assignment using a probability transformation.

B. Pignistic Probability Transform

The pignistic probability transformation ($BetP$) was first proposed by Philippe Smets in [22] and then included in [23] as a part of the Transferrable Belief Model. The pignistic probability transform involves transferring the belief mass from each non-singleton element of a BMA to its respective singleton elements by dividing its mass equally (i.e., according to its cardinality). The pignistic probability can be defined for any $X \subseteq \Omega$ as

$$BetP(X) = \sum_{\substack{Z \subseteq \Omega \\ Z \neq \emptyset}} \frac{|X \cap Z|}{|Z|} m(Z), \tag{9}$$

where $|\cdot|$ is the cardinality of a set. The pignistic probability transform satisfies all three Kolmogorov Axioms, and hence it only needs to be computed for the singleton elements $\omega \in \Omega$.

C. Sudano Proability Transforms

Sudano has proposed a suite of five probability transformations in [11]: $PrPl$, $PrNPl$, $PraPl$, $PrBel$, and $PrHyb$. In the current study, we will only focus on the $PrNPl$, $PrPl$, and $PrHyb$ transforms[1]. These three probability transforms are defined for the singleton elements $\omega \in \Omega$ as

$$PrPl(\omega) = \text{Pl}(\omega) \sum_{\substack{Z \subseteq \Omega \\ \omega \in Z}} \frac{m(Z)}{\sum_{\hat{\omega} \in Z} \text{Pl}(\hat{\omega})}, \tag{10}$$

$$PrNPl(\omega) = \frac{\text{Pl}(\omega)}{\sum_{\hat{\omega} \in \Omega} \text{Pl}(\hat{\omega})}, \tag{11}$$

and

$$PrHyb(\omega) = PraPl(\omega) \sum_{\substack{Z \subseteq \Omega \\ \omega \in Z}} \frac{m(Z)}{\sum_{\hat{\omega} \in Z} PraPl(\hat{\omega})}, \tag{12}$$

where $PraPl(\omega) = PrPl(\omega)$ for binary Ω. We direct the reader to [11] or [13] for the full definition of $PraPl(\omega)$.

D. Dezert-Smarandache's Probability Transform

A recently proposed probability transformation by Dezert and Smarandache, denoted $DSmP$, is introduced in [13]. $DSmP$ distributes belief masses assigned to the non-singleton elements of Ω proportionally, according to the belief masses assigned to the singleton elements. The transformation is defined in [13] using Dedekind lattices (i.e., the hyperpowerset of a set of alternatives). It is defined in terms of the powerset as

$$DSmP_\epsilon(X) = \sum_{Z \subseteq \Omega} \frac{\sum_{\hat{\omega} \in X \cap Z} m(\hat{\omega}) + \epsilon |X \cap Z|}{\sum_{\hat{\omega} \in Z} m(\hat{\omega}) + \epsilon |Z|} m(Z), \tag{13}$$

where $\epsilon \in [0, \infty]$ is a tuning parameter. As $\epsilon \to 0$, $DSmP$ approaches Sudano's $PrBel$ transform. As $\epsilon \to \infty$, $DSmP$ approaches $BetP$ [13]. The authors of [13] suggest selecting a small value for ϵ in order to minimize the amount of entropy present in the probabilities resulting from transformation. With this in mind, we used here $\epsilon = 0.001$. Similar to $BetP$, $DSmP$ satisfies all three Kolmogorov Axioms and only needs to be computed for the singleton elements $\omega \in \Omega$.

E. Probability Information Content

The probability information content (PIC) was proposed by Sudano in [14] as a method for comparing the performance of various probability transformations. For a subjective probability assignment $\mathcal{P}(\omega)$ genereated by the probability transformation $\mathcal{P}$, the PIC is defined as as

$$PIC_{\mathcal{P}(\omega)} = 1 + \frac{1}{\log M} \sum_{\omega \in \Omega} \mathcal{P}(\omega) \log \mathcal{P}(\omega) \tag{14}$$

where $M = |\Omega|$. A lower PIC value represents a subjective probability assignment where the alternatives are close to

[1]For binary decision tasks, $PrPl$ and $PraPl$ are equivalent [13]. Furthermore, the later defined $DSmP$ transform provides a more mathematically robust definition over $PrBel$ [13].

TABLE II
AVERAGE BMA AFTER COMBINATION (H_1 TRUE, LINE TASK).

Line Length Difference	$m(H_0)$	$m(H_1)$	$m(H_0 \cup H_1)$
0.27 mm	0.22	0.47	0.31
0.59 mm	0.08	0.69	0.23
1.23 mm	0.01	0.87	0.11
1.87 mm	0.00	0.94	0.05

TABLE III
AVERAGE PIC FOR EACH PROBABILITY TRANSFORMATION (H_1 TRUE, LINE TASK).

Line Length Difference	$BetP$	$PrPl$	$PrNPl$	$PrHyb$	$DSmP$
0.27 mm	0.56	0.63	0.51	0.68	0.83
0.59 mm	0.67	0.72	0.64	0.76	0.88
1.23 mm	0.85	0.87	0.84	0.89	0.95
1.87 mm	0.93	0.94	0.92	0.95	0.98

TABLE IV
AVERAGE BMA AFTER COMBINATION (H_1 TRUE, CITY TASK).

City Rank Difference	$m(H_0)$	$m(H_1)$	$m(H_0 \cup H_1)$
1 - 9	0.25	0.42	0.33
10 - 18	0.17	0.53	0.30
19 - 29	0.10	0.64	0.26
30 - 43	0.06	0.72	0.21

TABLE V
AVERAGE PIC FOR EACH PROBABILITY TRANSFORMATION (H_1 TRUE, CITY TASK).

City Rank Difference	$BetP$	$PrPl$	$PrNPl$	$PrHyb$	$DSmP$
1 - 9	0.49	0.57	0.43	0.64	0.80
10 - 18	0.53	0.61	0.48	0.67	0.82
19 - 29	0.60	0.67	0.56	0.73	0.85
30 - 43	0.68	0.73	0.64	0.78	0.89

being equiprobable. A higher PIC value represents a subjective probability assignment where one of the alternatives is close to having probability one.

IV. SIMULATION OVERVIEW

The 2DSD parameter sets relating to human responses on the line length discrimination task and the city population size discrimination task were used to simulate the decision performance of the five probability transformations of Section III (i.e., $BetP$, $PrPl$, $PrNPl$, $PrHyb$, and $DSmP$). Similar to what we have done in [17], a pool of 24 human responses were generated for each task using the parameter sets of the six subjects given in [15, Tables 3 and 6] by simulating four pairs of decisions and confidence assessments from each subject. Subjects were simulated over the first four difficulty levels given in Table I. For the line length discrimination task, we let H_1 denote the hypothesis that the second line presented to the subject is longer than the first and let H_0 denote the hypothesis that the first line presented to the subject is longer than the second. For the city population size discrimination task, we let H_1 denote the hypothesis that the second city presented to the subject has a higher population than the first and let H_0 denote the hypothesis that the first city presented has a higher population than the second. For each task, 10,000 trials were conducted for H_1 being true and 10,000 trials for H_0 being true. During each trial, the human responses from the subject pools were used to generate BMAs such that

$$m_i(X) = \begin{cases} p_i & X = a_i \\ 1 - p_i & X = \Omega \\ 0 & \text{otherwise} \end{cases}. \tag{15}$$

Here the subject decisions are given as $a_i \in \Omega = \{H_0, H_1\}$ and confidence assessments as $p_i \in [0, 1]$ for each subject $i = 1, \ldots, 24$. The subject BMAs m_i were combined two at a time using Yager's rule, as described by equations (7) and (8). Since Yager's rule is not associative, the combination order was randomized by choosing from the subject pool uniformly. The final combined BMAs were then each transformed into subjective probability assignments $\mathcal{P}_{1,\ldots,24}(\cdot)$ for each trial using the five probability transformations, $BetP$, $PrPl$, $PrNPl$, $PrHyb$, and $DSmP$.

Using $\mathcal{P}_{1,\ldots,24}(H_1)$, the fused decision $a_0 \in \{H_0, H_1\}$ was determined such that

$$a_0 = \begin{cases} H_1 & \mathcal{P}_{1,\ldots,24}(H_1) \geq \lambda \\ H_0 & \text{otherwise} \end{cases}, \tag{16}$$

where the threshold λ is varied in $[0, 1]$ for a desired detection and false alarm rate pair. The fused false alarm and detection rates are given as

$$FAR = P\left(\mathcal{P}_{1,\ldots,24}(H_1) \geq \lambda | H_0\right), \tag{17}$$

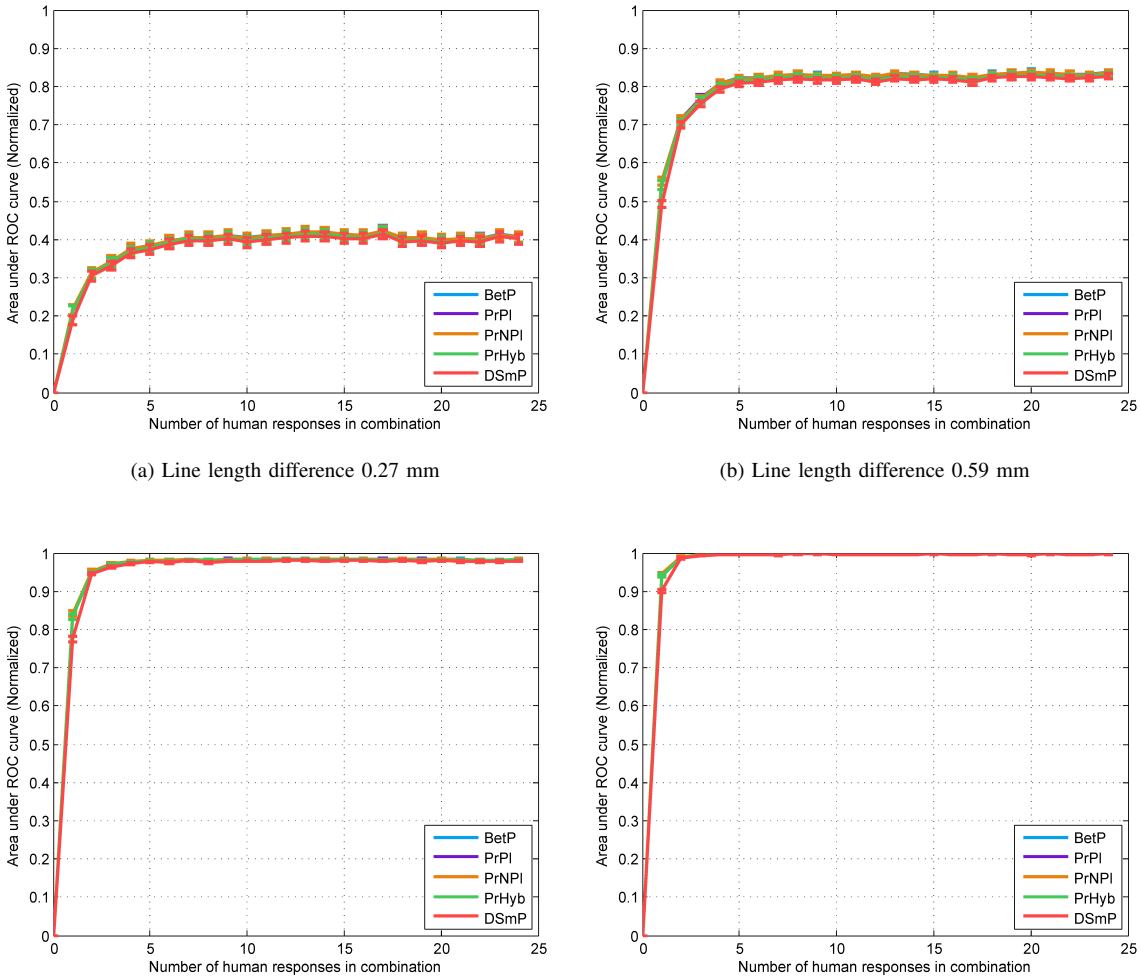
and

$$DET = P\left(\mathcal{P}_{1,\ldots,24}(H_1) \geq \lambda | H_1\right). \tag{18}$$

The threshold test of equation (16) was used to estimate false alarm rates and detection rates for varying threshold values λ using the 10,000 simulated responses with H_1 true and the 10,000 responses with H_0 true (after applying Yager's rule and each of the five probability transforms). These sets of false alarm and detection rates were then used to create ROC curves, and measure the areas under the ROC curves (AUCs) for each probability transform. Higher AUC values are indicative of higher discriminating performance between alternatives (i.e., higher detection rates for the same false alarm error rates). For comparison, the combined BMAs and the PIC values for each probability transformation were determined (averaged over 10,000 trials).

V. RESULTS

Tables II and IV show the average BMA for the line length discrimination task and the city population size discrimination task when H_1 is true (i.e., the second line is longer than the first, or the second city has a larger population than the first). The average BMAs for H_0 being true were the same, except that the values of $m(H_0)$ and $m(H_1)$ were reversed.



(a) Line length difference 0.27 mm

(b) Line length difference 0.59 mm

(c) Line length difference 1.23 mm

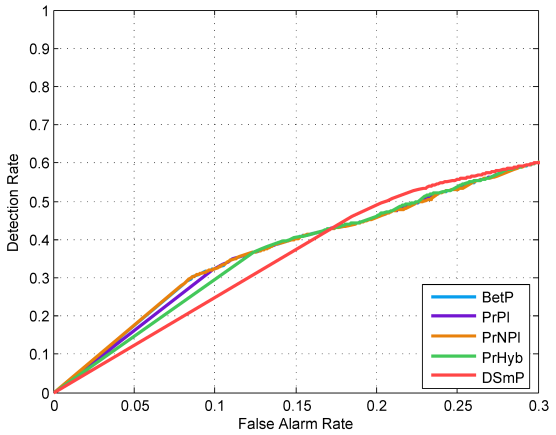
(d) Line length difference 1.87 mm

Fig. 1. Normalized area under the ROC curve (AUC) versus the number of sources present in combination, for each difficulty level of the line length discrimination task. Different lines represent the five different probability transforms investigated by this work. Error bars shown for the 95% confidence intervals. In each of the four difficulty levels, all five probability transforms are nearly overlapping.

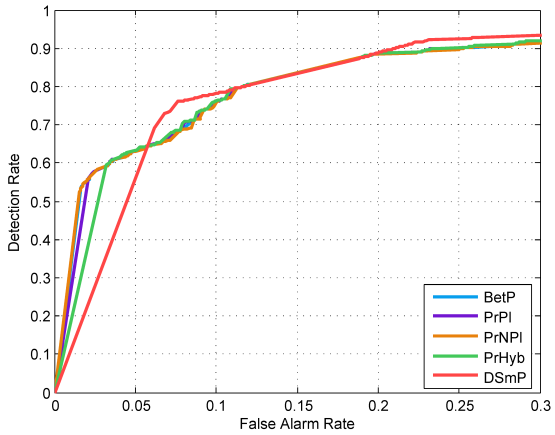
As expected, Yager's rule yielded combined BMAs which had high levels of imprecision (i.e., $m(H_0 \cup H_1)$ between 0.05 and 0.31 for the line length discrimination task and between 0.21 and 0.33 for the city population size discrimination task). As the task becomes easier (i.e., increasing the length difference between line pairs or increasing the population rank difference between city pairs), more belief mass is placed on the singleton elements. Tables III and V show the average PIC values after applying each of the five probability transformations to the combined BMAs resulting from the line length discrimination task and the city population size discrimination task. Higher PIC values are usually considered better [13], [14], [24]. The results in Tables III and V show increasing PIC values as task difficulty decreases, which seems reasonable. For each difficulty level, the PIC values follow the same trend with $PrNPl$ having the lowest PIC and $DSmP$ having the highest PIC. These trends supports the notion that DSmP produces subjective probabilities which are the most committed towards one of the alternatives (i.e., having the lowest entropy) [13].

Figure 1 shows the normalized AUCs versus the number of human responses present in combination for all five probability transforms on the line length discrimination task. Figure 3 shows the same quantities for the city population size discrimination task. Each subplot of figures 1 and 3 shows normalized AUCs for the four task difficulty levels simulated here. In all cases, the error bars represent the 95% confidence intervals. As expected, normalized AUC values increase as the task becomes easier. For any given difficulty level however, all five probability transformations exhibited statistically insignificant differences between normalized AUC values. The overall discriminating performance of all probability transforms is in fact the same.

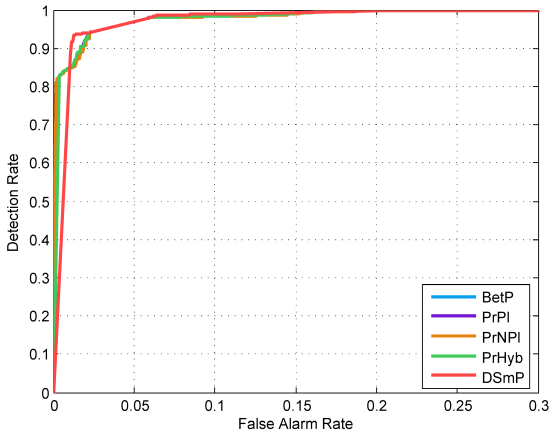
The ROC curves after combination for all five probability transforms are shown in Figure 2 for the line length



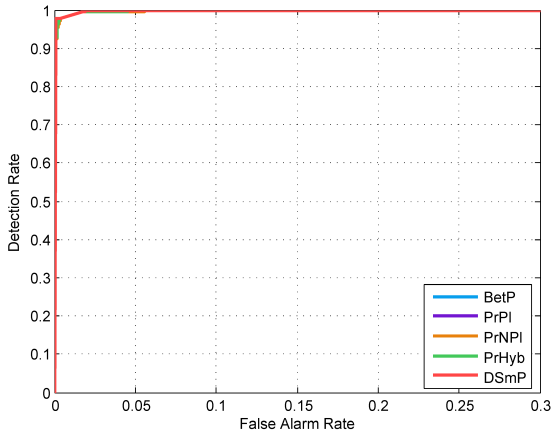
(a) Line length difference 0.27 mm



(b) Line length difference 0.59 mm



(c) Line length difference 1.23 mm



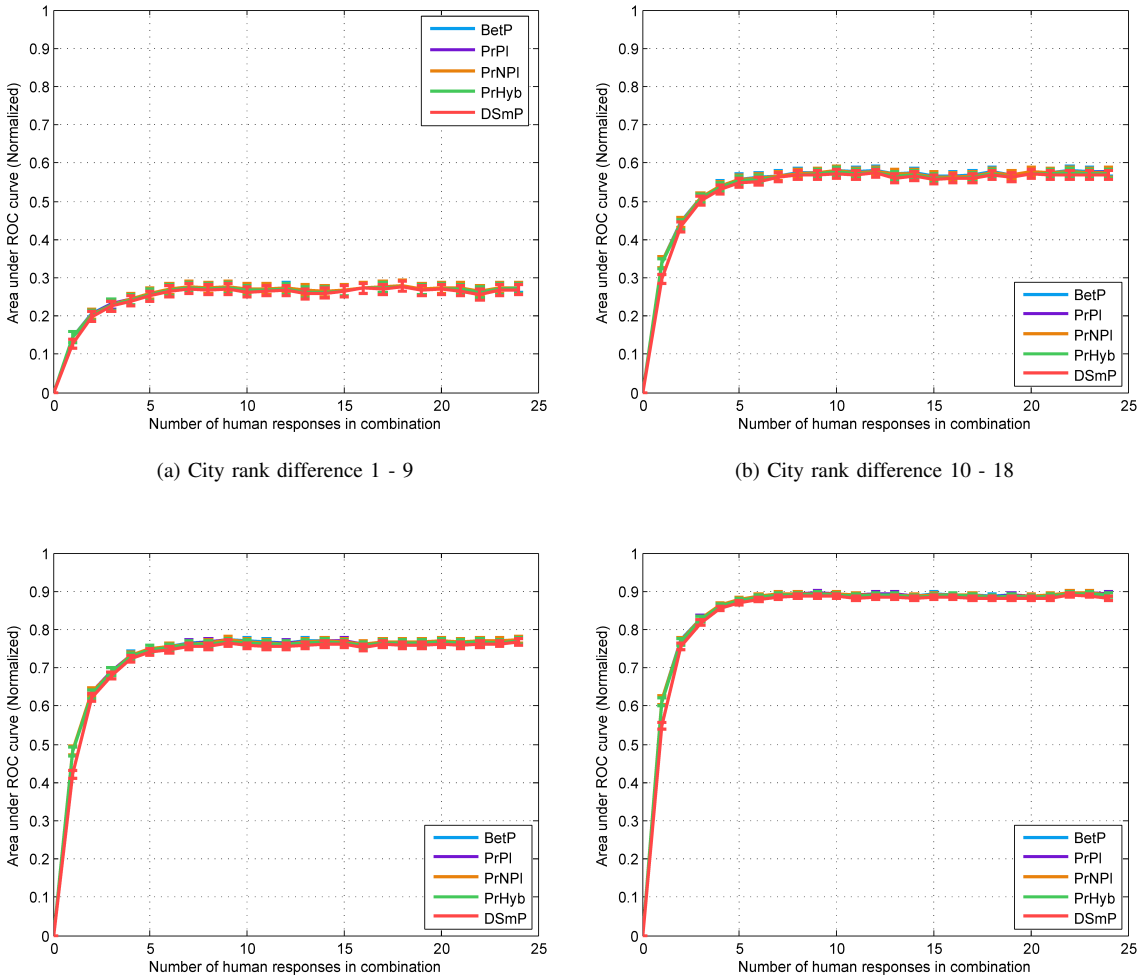
(d) Line length difference 1.87 mm

Fig. 2. ROC curves for each difficulty level of the line length discrimination task, showing false alarm rates up less than 0.30. Different lines represent the five different probability transforms investigated.

discrimination task and in Figure 4 for the city population size task. Each subplot of figures 2 and 4 shows ROCs for the four task difficulty levels. The ranges of the graph axes correspond to "reasonable" false alarm rates (i.e., up to 0.30). Again, the overall shape of the ROC for all probability transforms improves as the tasks become easier (supporting the results shown in Figure 3). For lower false alarm rates (e.g., less than 0.07), $BetP$, $PrPl$, $PrNPl$, and $PrHyb$ produce similar detection rates, which are all higher than those produced by $DSmP$. For higher false alarm rates (e.g., greater than 0.10), $DSmP$ produces higher detection rates over the remaining four probability transforms. As false alarm rates become even higher (e.g., greater than 0.25), similar performance is observed for all probability transforms. For the hardest and easiest variations of both tasks, these performance gains become less apparent. These observations support the conclusions reached by [25]; depending on the acceptable error rate for a specific task, a higher PIC value may not necessarily indicate higher detection rates.

VI. Conclusions

Using models of human responses from cognitive psychology, we have shown that there exist cases when the probability information content (PIC) does not always indicate "better" probability transforms. In the example presented here, two-stage dynamic signal detection (2DSD) was used to show that increasing PIC values may not necessarily lead to better discriminability performance of a probability transform. Specifically, it was found that all five probability transforms (i.e., $BetP$, $PrPl$, $PrNPl$, $PrHyb$, and $DSmP$) yielded the same discriminability performance (i.e., normalized AUC values) regardless of the number of sources included in the combination. Furthermore, the results indicate that some probability transforms yield higher detection rates over others, depending on the false alarm rate required. For lower false alarm rates (e.g., less than 0.07), the $BetP$, $PrPl$, $PrNPl$,



(a) City rank difference 1 - 9

(b) City rank difference 10 - 18

(c) City rank difference 19 - 29

(d) City rank difference 30 - 43

Fig. 3. Normalized area under the ROC curve (AUC) versus the number of sources present in combination, for each difficulty level of the city population size discrimination task. Different lines represent the five different probability transforms investigated by this work. Error bars shown for the 95% confidence intervals. In each of the four difficulty levels, all five probability transforms are nearly overlapping.

and $PrHyb$ transforms yielded higher detection rates than $DSmP$. For higher false alarm rates (e.g., greater than 0.10), this trend was reversed. These findings support the arguments presented in [25], and suggest that simulation and testing should be performed before the components of a specific fusion system are selected.

REFERENCES

[1] T. Wickramarathne, K. Premaratne, M. Murthi, M. Scheutz, S. Kubler, and M. Pravia, "Belief theoretic methods for soft and hard data fusion," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 2388–2391.

[2] A. Moro, E. Mumolo, M. Nolich, K. Terabayashi, and K. Umeda, "Improved foreground-background segmentation using dempster-shafer fusion," in *8th International Symposium on Image and Signal Processing and Analysis*, 2013, pp. 72–77.

[3] K. Premaratne, M. Murthi, J. Zhang, M. Scheutz, and P. Bauer, "A dempster-shafer theoretic conditional approach to evidence updating for fusion of hard and soft data," in *12th International Conference on Information Fusion*, July 2009, pp. 2122–2129.

[4] J. Li, Y. Wang, and Z. J. Mao, "A new multisensor information fusion model using dempster-shafer theory," *Applied Mechanics and Materials*, vol. 475, pp. 415–418, 2014.

[5] O. Basir and X. Yuan, "Engine fault diagnosis based on multi-sensor information fusion using dempster–shafer evidence theory," *Information Fusion*, vol. 8, no. 4, pp. 379–386, 2007.

[6] S. Acharya and M. Kam, "Evidence combination for hard and soft sensor data fusion," in *Proceedings of the 14th International Conference on Information Fusion*, July 2011, pp. 1–8.

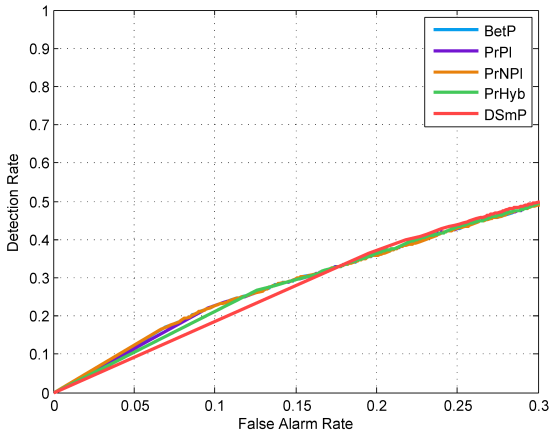
[7] W. Talbott, "Bayesian epistemology," in *The Stanford Encyclopedia of Philosophy*, E. N. Zalta, Ed. Stanford University, 2011. [Online]. Available: http://plato.stanford.edu/archives/sum2011/entries/epistemology-bayesian/

[8] G. Shafer, *A Mathematical Theory of Evidence*. Princeton University Press: Princeton, NJ, 1976.

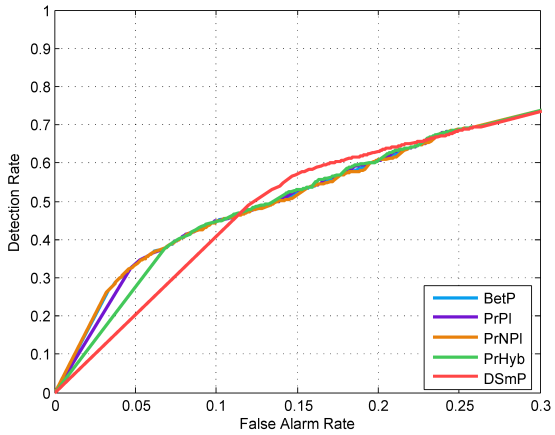
[9] J. Pearl, "Reasoning with belief functions: An analysis of compatibility," *International Journal of Approximate Reasoning*, vol. 4, no. 5-6, pp. 363–389, 1990.

[10] P. Smets, "Analyzing the combination of conflicting belief functions," *Information Fusion*, vol. 8, no. 4, pp. 387–412, 2007.

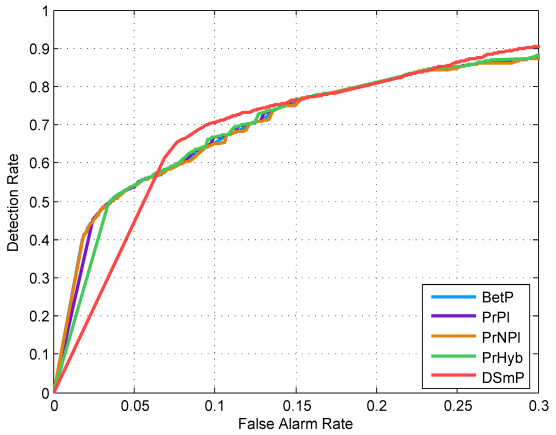
[11] J. Sudano, "Yet another paradigm illustrating evidence fusion," in *Proceedings of the 9th International Conference on Information Fusion*,



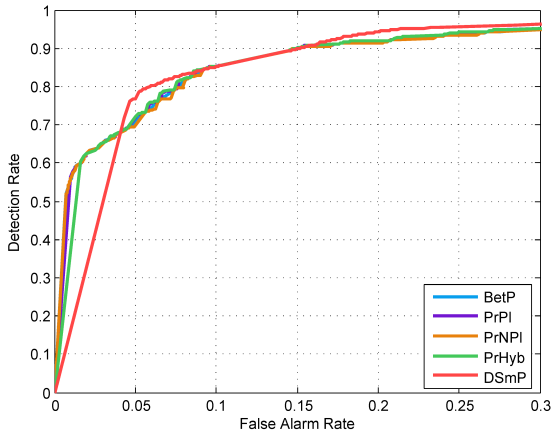
(a) City rank difference 1 - 9



(b) City rank difference 10 - 18



(c) City rank difference 19 - 29



(d) City rank difference 30 - 43

Fig. 4. ROC curves for each difficulty level of the city population size discrimination task, showing false alarm rates up less than 0.30. Different lines represent the five different probability transforms investigated.

2006.

[12] F. Cuzzolin, "A geometric approach to the theory of evidence," *IEEE Transactions on Systems Man and Cybernetics, Part C*, vol. 38, p. 2008, 2008.

[13] J. Dezert and F. Smarandache, "A new probabilistic transformation of belief mass assignment," in *Proceedings of the 11th International Conference on Information Fusion*, 2008, pp. 1–8.

[14] J. Sudano, "Pignistic probability transforms for mixes of low-and high-probability events," in *4th International Conference on Information Fusion*, 2001.

[15] T. Pleskac and J. Busemeyer, "Two-stage dynamic signal detection: A theory of choice, decision time, and confidence," *Psychological review*, vol. 117, no. 3, pp. 864–901, July 2010.

[16] D. J. Bucci, S. Acharya, and M. Kam, "Simulating human decision making for testing soft and hard/soft fusion algorithms," in *Proceedings of the 47th Annual Conference on Information Sciences and Systems (CISS)*, 2013.

[17] D. J. Bucci, S. Acharya, T. J. Pleskac, and M. Kam, "Subjective confidence and source reliability in soft data fusion," in *Proceedings of the 48th Annual Conference on Information Sciences and Systems (CISS)*, March 2014.

[18] R. Yager, "On the dempster-shafer framework and new combination rules," *Information Sciences*, vol. 41, pp. 93–138, 1987.

[19] P. Smets, "Decision making in the tbm: The necessity of the pignistic transformation," *International Journal of Approximate Reasoning*, vol. 38, pp. 133–147, 2005.

[20] A. Heathcote, S. Brown, and D. Mewhort, "Quantile maximum likelihood estimation of response time distributions," *Psychonomic Bulletin & Review*, vol. 9, no. 2, pp. 394–401, 2002.

[21] United States Census Bureau. Vintage 2006: Metropolitan and micropolitan statistical areas tables. [Online]. Available: http://www.census.gov/popest/data/historical/2000s/vintage_2006/metro.html

[22] P. Smets, "Constructing the pignistic probability function in a context of uncertainty," *Uncertainty in AI*, vol. 5, pp. 29–39, 1990.

[23] P. Smets and R. Kennes, "The transferrable belief model," *Artificial Intelligence*, vol. 66, pp. 191–234, 1994.

[24] J. J. Sudano, "The system probability information content (pic) relationship to contributing components, combining independent multi-source beliefs, hybrid and pedigree pignistic probabilities," in *Proceedings of the 5th International Conference on Information Fusion*. IEEE, 2002, pp. 1277–1283.

[25] D. Han, J. Dezert, C. Han, and Y. Yang, "Is entropy enough to evaluate the probability transformation approach of belief function?" in *Proceedings of the 13th Conference on Information Fusion*, 2010, pp. 1–7.