MINIMUM MEAN BAYES RISK ERROR QUANTIZATION OF PRIOR PROBABILITIES

Kush R. Varshney and Lav R. Varshney

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology, Cambridge, MA, USA

ABSTRACT

Bayesian hypothesis testing is investigated when the prior probabilities of the hypotheses, taken as a random vector, must be quantized. Nearest neighbor and centroid conditions for quantizer optimality are derived using mean Bayes risk error as a distortion measure. An example of optimal quantization for hypothesis testing is provided. Human decision making is briefly studied assuming quantized prior Bayesian hypothesis testing; this model explains several experimental findings.

Index Terms— quantization, categorization, Bayesian hypothesis testing, signal detection, Bayes risk error

1. INTRODUCTION

Consider a hypothesis testing scenario in which an object is to be observed to determine which one of M states, $\{h_0, \ldots, h_{M-1}\}$, it is in. The object has prior probability p_m of being in state m, i.e. $p_m = \Pr[H = h_m]$, and prior probability vector $\boldsymbol{p} = \begin{bmatrix} p_0 & \cdots & p_{M-1} \end{bmatrix}^T$, with $\sum_{m=0}^{M-1} p_m = 1$, which is known to the decision maker. M-ary hypothesis testing with known prior probabilities calls for the Bayesian formulation to the problem, for which the optimal decision rule minimizes Bayes risk.

Now consider the situation when there is a population of objects, each with its own prior probability vector drawn from the distribution $f_{\boldsymbol{P}}(\boldsymbol{p})$ supported on the M-dimensional probability simplex. If the prior probability vector of each object were known perfectly to the decision maker before observation and hypothesis testing, then the scenario would be no different than that of standard Bayesian hypothesis testing. However, we consider the case in which the decision maker is constrained and can only work with at most K different prior probability vectors. Hence, when there are more than K objects, the decision maker must first map the true prior probability vector of the object being observed to one of the K available vectors and then proceed to perform the optimal Bayesian hypothesis test, treating that vector as the prior probabilities of the object. The decision maker performs the mapping operation without error.

In this paper, the design of the mapping from prior probability vectors in the population to one of K representative probability vectors is approached as a quantization problem. Mean Bayes risk error (MBRE) is defined as a fidelity criterion for the quantization of $f_{\boldsymbol{P}}(\boldsymbol{p})$ and conditions are derived for a minimum MBRE quantizer. Some examples of MBRE-optimal quantizers are given along with their performance in the low-rate quantization regime. We also discuss how certain tasks that human decision makers face are well-modeled by the hypothesis testing scenario of this paper due to certain suboptimalities in human processing.

This work was supported in part by NSF Graduate Research Fellowships, and MURIs funded through ARO Grant W911NF-06-1-0076 and through AFOSR Grant FA9550-06-1-0324.

Note that previous work that combines detection and quantization looks at the quantization of observed data, not prior probabilities, and also only approximates the Bayes risk function instead of working with it directly [1, 2, 3]. In such work, the concern is the scenario in which there is a communication constraint between the sensor and the decision maker, but the decision maker has unconstrained processing capability. We are concerned with the opposite case, in which there is no communication constraint between the sensor and decision maker, but the decision maker must operate under a finite memory constraint. Finite memory constraints apply not just to humans but to all decision making systems. We are not aware of any previous work that has looked at quantization, clustering, or categorization of prior probabilities. In the remainder of the paper, we focus on binary hypothesis testing, $M = 2$.

2. BAYES RISK ERROR

In the binary Bayesian hypothesis testing problem, there are two hypotheses h_0 and h_1 with prior probabilities $p_0 = \Pr[H = h_0]$ and $p_1 = \Pr[H = h_1] = 1 - p_0$, a noisy observation Y, and likelihoods $f_{Y|H}(y|h_0)$ and $f_{Y|H}(y|h_1)$. A function $\hat{h}(y)$ is designed that uniquely maps every possible y to either h_0 or h_1. The two types of error probabilities are $p_E^{\mathrm{I}} = \Pr[\hat{h}(Y) = h_1 | H = h_0]$ and $p_E^{\mathrm{II}} = \Pr[\hat{h}(Y) = h_0 | H = h_1]$.

The decision rule $\hat{h}(y)$ is chosen to minimize the Bayes risk function J, an expectation over the non-negative cost function $c(h_i, h_j)$:

$$\begin{aligned} J &= E[c(H, \hat{h}(Y))] \\ &= (c_{10} - c_{00}) p_0 p_E^{\mathrm{I}} + (c_{01} - c_{11}) p_1 p_E^{\mathrm{II}} + c_{00} p_0 + c_{11} p_1, \end{aligned} \tag{1}$$

where $c_{ij} = c(h_i, h_j)$. It is often of interest to assign no cost to correct decisions, i.e. $c_{00} = c_{11} = 0$, which we assume in the remainder of this paper. In this case, the Bayes risk simplifies to:

$$J(p_0) = c_{10} p_0 p_E^{\mathrm{I}}(p_0) + c_{01}(1 - p_0) p_E^{\mathrm{II}}(p_0). \tag{2}$$

In (2), the dependence of the Bayes risk and error probabilities on p_0 has been explicitly noted (p_1 is automatically specified by specifying p_0). The function $J(p_0)$ is zero at the points $p_0 = 0$ and $p_0 = 1$ and is positive-valued, concave, and continuous in the interval $(0, 1)$.

In the case when the true prior probability is p_0, but $\hat{h}(y)$ is designed using some other value a, there is mismatch, and the mismatched Bayes risk is:

$$\tilde{J}(p_0, a) = c_{10} p_0 p_E^{\mathrm{I}}(a) + c_{01}(1 - p_0) p_E^{\mathrm{II}}(a). \tag{3}$$

$\tilde{J}(p_0, a)$ is a linear function of p_0 with slope $(c_{10} p_E^{\mathrm{I}}(a) - c_{01} p_E^{\mathrm{II}}(a))$ and intercept $c_{01} p_E^{\mathrm{II}}(a)$. Note that $\tilde{J}(p_0, a)$ is tangent to $J(p_0)$ at a and that $\tilde{J}(p_0, p_0) = J(p_0)$.

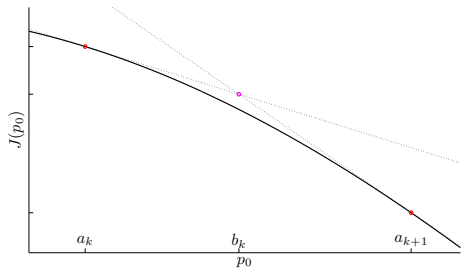

Fig. 1. The intersection of the lines $\tilde{J}(p_0, a_k)$, tangent to $J(p_0)$ at a_k, and $\tilde{J}(p_0, a_{k+1})$, tangent to $J(p_0)$ at a_{k+1}, is the optimal interval boundary.

Let us define Bayes risk error as the difference between the mismatched Bayes risk function and the Bayes risk function:

$$d(p_0, a) = \tilde{J}(p_0, a) - J(p_0). \tag{4}$$

Since $J(p_0)$ is a non-negative continuous, concave function and the line $\tilde{J}(p_0, a)$ is tangent to $J(p_0)$, we know that $\tilde{J}(p_0, a) \geq J(p_0) \geq 0$. Consequently, $d(p_0, a)$ is non-negative and only equal to zero when $p_0 = a$. Moreover, $d(p_0, a)$ is convex in both p_0 and a and continuous in p_0 for all a.

3. QUANTIZER OPTIMALITY CONDITIONS

The conditions necessary for the local optimality of a scalar quantizer for $f_{P_0}(p_0)$ under Bayes risk error distortion are now derived. A K-point scalar quantizer partitions the interval $[0, 1]$ into K subintervals $\mathcal{R}_1 = [0, b_1]$, $\mathcal{R}_2 = (b_1, b_2]$, $\mathcal{R}_3 = (b_2, b_3]$, $\ldots$, $\mathcal{R}_K = (b_{K-1}, 1]$. For each of these quantization regions $\mathcal{R}_k$, there is a representation point or codeword, a_k, to which elements are mapped. A quantizer can be viewed as a nonlinear function $v(\cdot)$ such that $v(p_0) = a_k$ for $p_0 \in \mathcal{R}_k$. For a given K, we would like to find the quantizer that minimizes the MBRE:

$$D = E[d(P_0, v(P_0))] = \int d(p_0, v(p_0)) f_{P_0}(p_0) dp_0. \tag{5}$$

There is no closed-form solution, but an optimal quantizer must satisfy the nearest neighbor condition, the centroid condition, and the zero probability of boundary condition [4]. The nearest neighbor and centroid conditions are developed for MBRE in the following subsections. Using the nearest neighbor and centroid conditions, the iterative Lloyd-Max algorithm can be applied to find minimum MBRE quantizers [4].

3.1. Nearest Neighbor Condition

With the codebook $\{a_k\}$ fixed, an expression for the interval boundaries $\{b_k\}$ is derived. Given any $p_0 \in [a_k, a_{k+1}]$, if $\tilde{J}(p_0, a_k) < \tilde{J}(p_0, a_{k+1})$ then Bayes risk error is minimized if p_0 is represented by a_k and vice versa. The boundary point $b_k \in [a_k, a_{k+1}]$ is the abscissa of the point at which the lines $\tilde{J}(p_0, a_k)$ and $\tilde{J}(p_0, a_{k+1})$ intersect. The idea is illustrated graphically in Fig. 1.

By manipulating the slopes and intercepts of $\tilde{J}(p_0, a_k)$ and $\tilde{J}(p_0, a_{k+1})$, the point of intersection is found to be:

$$b_k = \frac{c_{01}\left(p_E^{\mathrm{II}}(a_{k+1}) - p_E^{\mathrm{II}}(a_k)\right)}{c_{01}\left(p_E^{\mathrm{II}}(a_{k+1}) - p_E^{\mathrm{II}}(a_k)\right) - c_{10}\left(p_E^{\mathrm{I}}(a_{k+1}) - p_E^{\mathrm{I}}(a_k)\right)}. \tag{6}$$

3.2. Centroid Condition

With the quantization regions fixed, the MBRE is to be minimized over the $\{a_k\}$. Here, the MBRE is expressed as the sum of integrals over quantization regions:

$$D = \sum_{k=1}^{K} \int_{\mathcal{R}_k} \left(\tilde{J}(p_0, a_k) - J(p_0)\right) f_{P_0}(p_0) dp_0. \tag{7}$$

Because the regions are fixed, the minimization may be performed for each interval separately.

Let us define $I_k^{\mathrm{I}} = \int_{\mathcal{R}_k} p_0 f_{P_0}(p_0) dp_0$ and $I_k^{\mathrm{II}} = \int_{\mathcal{R}_k} (1 - p_0) f_{P_0}(p_0) dp_0$, which are conditional means. Then:

$$a_k = \arg\min_a \left\{ c_{10} I_k^{\mathrm{I}} p_E^{\mathrm{I}}(a) + c_{01} I_k^{\mathrm{II}} p_E^{\mathrm{II}}(a) \right\}. \tag{8}$$

Since $d(p_0, a)$ is convex, (8) is uniquely minimized by setting its derivative equal to zero. Thus, a_k is the solution to:

$$c_{10} I_k^{\mathrm{I}} \tfrac{dp_E^{\mathrm{I}}(a_k)}{da_k} + c_{01} I_k^{\mathrm{II}} \tfrac{dp_E^{\mathrm{II}}(a_k)}{da_k} = 0. \tag{9}$$

4. EXAMPLES

As an example, let us consider the following scalar signal and measurement model:

$$Y = s_m + W, \quad m \in \{0, 1\}, \tag{10}$$

where $s_0 = 0$ and $s_1 = \mu$ (a known, deterministic quantity), and W is a zero-mean, Gaussian random variable with variance σ^2. The two error probabilities are:

$$\begin{aligned} p_E^{\mathrm{I}}(p_0) &= Q\left(\tfrac{\mu}{2\sigma} + \tfrac{\sigma}{\mu} \ln\left(\tfrac{c_{10} p_0}{c_{01}(1-p_0)}\right)\right), \\ p_E^{\mathrm{II}}(p_0) &= Q\left(\tfrac{\mu}{2\sigma} - \tfrac{\sigma}{\mu} \ln\left(\tfrac{c_{10} p_0}{c_{01}(1-p_0)}\right)\right), \end{aligned} \tag{11}$$

where $Q(\alpha) = \frac{1}{\sqrt{2\pi}} \int_\alpha^\infty e^{-x^2/2} dx$.

Finding the centroid condition, the derivatives of the error probabilities are:

$$\begin{aligned} \left.\tfrac{dp_E^{\mathrm{I}}(p_0)}{dp_0}\right|_{p_0 = a_k} &= -\tfrac{1}{\sqrt{2\pi}} \tfrac{\sigma}{\mu} \tfrac{1}{a_k(1-a_k)} e^{-\frac{1}{2}\left(\frac{\mu}{2\sigma} + \frac{\sigma}{\mu} \ln\left(\frac{c_{10} a_k}{c_{01}(1-a_k)}\right)\right)^2}, \\ \left.\tfrac{dp_E^{\mathrm{II}}(p_0)}{dp_0}\right|_{p_0 = a_k} &= +\tfrac{1}{\sqrt{2\pi}} \tfrac{\sigma}{\mu} \tfrac{1}{a_k(1-a_k)} e^{-\frac{1}{2}\left(\frac{\mu}{2\sigma} - \frac{\sigma}{\mu} \ln\left(\frac{c_{10} a_k}{c_{01}(1-a_k)}\right)\right)^2}. \end{aligned}$$

By substituting these derivatives into (9) and simplifying, the following expression is obtained for the representation points:

$$a_k = \frac{I_k^{\mathrm{I}}}{I_k^{\mathrm{I}} + I_k^{\mathrm{II}}}. \tag{12}$$

Examples with $\mu = 1$, $\sigma = 1$ are presented below. We look at the setting in which all prior probabilities are equally likely. As a point of reference, a comparison is made to quantizers designed under mean absolute error (MAE), an objective that does not account

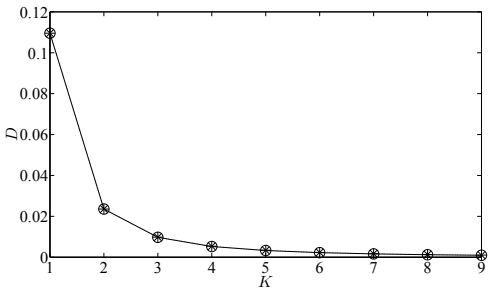

Fig. 2. MBRE for uniformly distributed P_0 and Bayes costs $c_{10} = c_{01} = 1$ plotted as a function of the number of quantization levels K; the solid line with circle markers is the MBRE-optimal quantizer and the dotted line with asterisk markers is the MAE-optimal uniform quantizer. (The two lines are nearly coincident.)

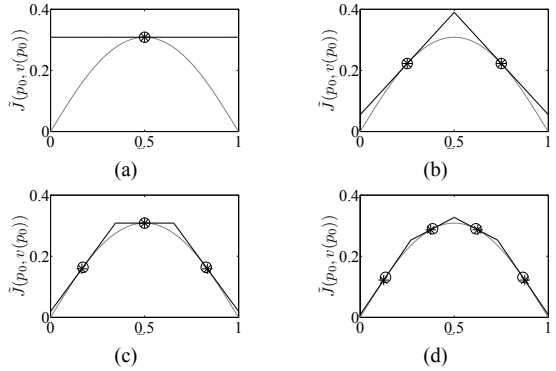

Fig. 3. Quantizers for uniformly distributed P_0 and Bayes costs $c_{10} = c_{01} = 1$. $\tilde{J}(p_0, v(p_0))$ is plotted for (a) $K = 1$, (b) $K = 2$, (c) $K = 3$, and (d) $K = 4$; the markers, circle and asterisk for the MBRE-optimal and MAE-optimal quantizers respectively, are the representation ponts $\{a_k\}$. The gray line is the unquantized Bayes risk $J(p_0)$.

for hypothesis testing [5]. The MBRE-optimal quantizer's MBRE is of course never worse than that of the MAE-optimal quantizer.

The MBRE of the MBRE-optimal quantizer and a quantizer designed to minimize MAE with respect to uniform $f_{P_0}(p_0)$ are plotted in Fig. 2. (The optimal MAE quantizer for the uniform distribution is the uniform quantizer.) The plot shows MBRE as a function of K; the solid line with circle markers is the MBRE-optimal quantizer and the dotted line with asterisk markers is the MAE-optimal quantizer.

The performance of both quantizers is similar, but the MBRE-optimal quantizer always performs better or equally. For $K = 1, 2$, the two quantizers are identical, as seen in Fig. 3a-b. The plots in Fig. 3 show $\tilde{J}(p_0, v(p_0))$ solid with circle markers and dotted with asterisk markers for the MBRE- and MAE-optimal quantizers respectively; the markers are the representation points. The gray line is $J(p_0)$, the Bayes risk with unquantized prior probabilities. Each increment of K is associated with a reduction in Bayes risk. There is a large performance improvement from $K = 1$ to $K = 2$.

In Fig. 4 and Fig. 5, similar plots to those above are given for the case when the Bayes costs c_{10} and c_{01} are unequal. The unequal

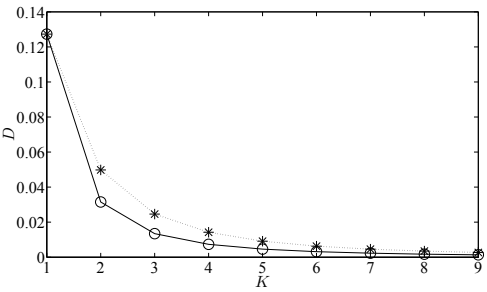

Fig. 4. MBRE for uniformly distributed P_0 and Bayes costs $c_{10} = 1, c_{01} = 4$ plotted as a function of the number of quantization levels K; the solid line with circle markers is the MBRE-optimal quantizer and the dotted line with asterisk markers is the MAE-optimal uniform quantizer.

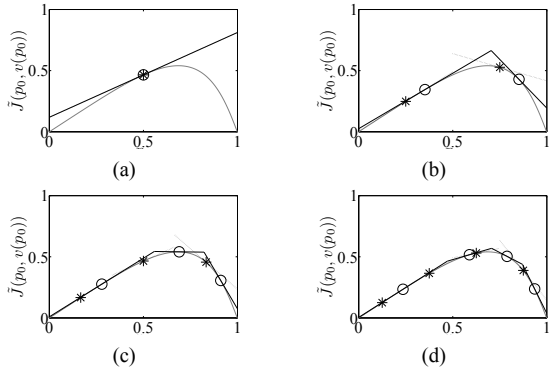

Fig. 5. Quantizers for uniformly distributed P_0 and Bayes costs $c_{10} = 1, c_{01} = 4$. $\tilde{J}(p_0, v(p_0))$ is plotted for (a) $K = 1$, (b) $K = 2$, (c) $K = 3$, and (d) $K = 4$; the markers, circle and asterisk for the MBRE-optimal and MAE-optimal quantizers respectively, are the representation ponts $\{a_k\}$. The gray line is the unquantized Bayes risk $J(p_0)$.

costs skew the Bayes risk function and consequently the representation point locations. The difference in performance between the MBRE-optimal and MAE-optimal quantizers is greater in this example because the MAE-criterion cannot incorporate the Bayes costs, which factor into MBRE calculation.

5. IMPLICATIONS ON HUMAN DECISION MAKING

Let us consider one particular setting for human decision making: a referee deciding whether a player has committed a foul using his or her noisy observation as well as prior experience. Every player commits fouls at a different rate; some players are dirtier or more agressive than others. It is this rate which is the prior probability for the 'foul committed' hypothesis. Hence, over the population of players, there is a distribution of prior probabilities. If the referee tunes the prior probability to the particular player on whose action the decision is to be made, decision-making performance is improved.

Human decision makers are limited in their information processing capacity and can only carry around seven, plus or minus two,

categories without getting confused [6]. Consequently, the referee is limited to categorizing players into a small number of dirtiness levels, with associated prototypical prior probabilities. This amounts to quantization of the distribution of prior probabilities and the use of the quantization level centroid in which a player falls as the prior probability when performing hypothesis testing on that player's action, exactly the scenario discussed in the previous sections.

A referee will do a better job with more categories rather than fewer. Implications of this sort are not surprising. However, when one additional component is added to the decision-making scenario, some fairly interesting implications arise.

We discuss mathematically unavoidable consequences of quantized prior hypothesis testing when quantizing the prior probability for a minority population and for a majority population separately, while taking identical prior probability distributions of the two populations $f_{P_0}(p_0)$. Distinct populations can be defined along any socially observable dimension; for ease of exposition we use 'white' and 'black' to denote the two populations. Although there is some debate in the social cognition literature [7], it is thought that race and gender categorization is essentially automatic, particularly when a perceiver lacks the motivation, time, or cognitive capacity to think deeply.

We can extend the definition of MBRE to two populations as:

$$D^{(2)} = \tfrac{w}{w+b} E[\tilde{J}(P_0, v_{K_w}(P_0))] + \tfrac{b}{w+b} E[\tilde{J}(P_0, v_{K_b}(P_0))] - E[J(P_0)], \quad (13)$$

where w is the number of whites encountered, b is the number of blacks encountered, K_w is the number of points in the quantizer for whites, and K_b is the number of points in the quantizer for blacks. In order to find the optimal allocation of the total quota of representation points $K_t = K_w + K_b$, we minimize $D^{(2)}$ for all $K_t - 1$ possible allocations and choose the best one.

Fryer and Jackson have previously suggested that it is better to allocate more representation points to the majority population than to the minority population [8]. With two separate scalar quantizers, but a single codebook size constraint, optimizing $D^{(2)}$ over $v_{K_w}(\cdot)$ and $v_{K_b}(\cdot)$ yields the same result. The MBRE for members of the minority group is greater than that for the majority group.

Assuming white decision makers have $w > b$ and black decision makers have $b > w$ due to different exposure [9], analysis of quantized prior Bayesian hypothesis testing predicts that there should be own-race bias in decision making. This prediction is in fact born out experimentally, see e.g. [10], and in data collected for econometric studies, e.g. [11, 12]. The human angle to hypothesis testing with quantized priors, including the role of the Bayes costs c_{10} and c_{01}, is discussed in greater detail in a manuscript by the authors [13].

6. CONCLUSION

We have looked at Bayesian hypothesis testing when there is a distribution of prior probabilities, but the decision maker may only use a quantized version of the true prior probability in designing a decision rule. Considering the problem of finding the optimal quantizer for this purpose, we have defined a new fidelity criterion based on the Bayes risk function. For this criterion, MBRE, we have determined the conditions that an optimal quantizer satisfies. M-ary hypothesis testing with $M > 2$ requires vector quantization rather than scalar quantization, but determining the Lloyd-Max conditions is no different conceptually due to the geometry of the Bayes risk function and mismatched Bayes risk function. For the M-ary hypothesis testing case, a multivariate distribution such as the M-dimensional Dirichlet distribution is needed for $f_{\boldsymbol{P}}(\boldsymbol{p})$. Previous, though significantly different, work on quantization for hypothesis testing was unable to directly minimize the Bayes risk, as was accomplished in this work.

Discrimination on the basis of race, gender, and other socially observable characteristics has been a troublesome social problem. Here we have formulated a mathematical theory of quantized prior hypothesis testing, which, when combined with theories of social cognition and empirical facts about segregation leads to a generative model of such discriminative behavior. This biased decision making arises despite having identical distributions for different populations and despite no malicious intent on the part of the decision maker. Discrimination appears to be a permanent artifact of the automaticity of classification along social dimensions and the finite human capacity for information processing.

7. ACKNOWLEDGMENT

The authors thank Vivek K Goyal, Sanjoy K. Mitter, and Alan S. Willsky.

8. REFERENCES

[1] S. A. Kassam, "Optimum quantization for signal detection," *IEEE Trans. Commun.*, vol. COM-25, pp. 479–484, May 1977.

[2] H. V. Poor and J. B. Thomas, "Applications of Ali-Silvey distance measures in the design of generalized quantizers," *IEEE Trans. Commun.*, vol. COM-25, pp. 893–900, Sept. 1977.

[3] R. Gupta and A. O. Hero, III, "High-rate vector quantization for detection," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1951–1969, Aug. 2003.

[4] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, Boston, 1992.

[5] S. A. Kassam, "Quantization based on the mean-absolute-error criterion," *IEEE Trans. Commun.*, vol. COM-26, pp. 267–270, Feb. 1978.

[6] G. A. Miller, "The magical number seven, plus or minus two: Some limits on our capacity for processing information," *Psychol. Rev.*, vol. 63, pp. 81–97, 1956.

[7] C. N. Macrae and G. V. Bodenhausen, "Social cognition: Thinking categorically about others," *Annu. Rev. Psychol.*, vol. 51, pp. 93–120, Feb. 2000.

[8] R. G. Fryer, Jr. and M. O. Jackson, "A categorical model of cognition and biased decision-making," *B. E. J. Theor. Econ.*, to appear.

[9] F. Echenique and R. G. Fryer, Jr., "A measure of segregation based on social interactions," *Quart. J. Econ.*, vol. 122, pp. 441–485, May 2007.

[10] C. A. Meissner and J. C. Brigham, "Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review," *Psychol. Pub. Pol. L.*, vol. 7, pp. 3–35, Jan. 2001.

[11] J. J. Donohue, III and S. D. Levitt, "The impact of race on policing and arrests," *J. Law Econ.*, vol. 44, pp. 367–394, Oct. 2001.

[12] J. Price and J. Wolfers, "Racial discrimination among NBA referees," Working Paper 13206, NBER, June 2007.

[13] L. R. Varshney and K. R. Varshney, "Bayesian hypothesis testing with prototype priors and its implications on social discrimination," Sept. 2007.