Towards a Generalized Bayesian Model of Category Effects

Zihao Xu[1], Pernille Hemmer[2,3], and Qiong Zhang[1,2,3]

Computer Science Department[1], Psychology Department[2], Center for Cognitive Science[3],

Rutgers University–New Brunswick

Recall the size of the apple

2.7 inches?

People's estimation is biased towards apple's average size.

Category Effect

An individual stimulus from a category is often judged to be closer to the center of that category than its true location.

2.7 inches?

Bartlett, 1932; Etcoff and Magee, 1992; Harnad, 1987; Huttenlocher et al., 2000; Liberman et al., 1957; Mitterer and De Ruiter, 2008; Persaud and Hemmer, 2014

Category Adjustment Model (CAM) (Huttenlocher et al., 2000)

An individual stimulus from a category is often judged to be closer to the center of that category than its true location.

$$\hat{T} = w\mu_c + (1 - w)M$$

$\hat{T}$: Human Recall, μ_c: Category Center, M: Noisy Memory, w: Constant

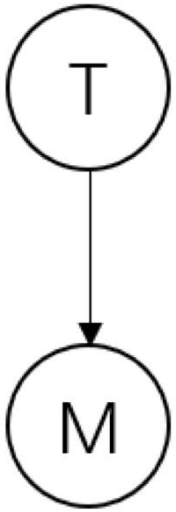
Category Adjustment Model (CAM) (Huttenlocher et al., 2000)

True stimulus T is drawn from a category c. $T \sim N(\mu_c, \sigma_c^2)$

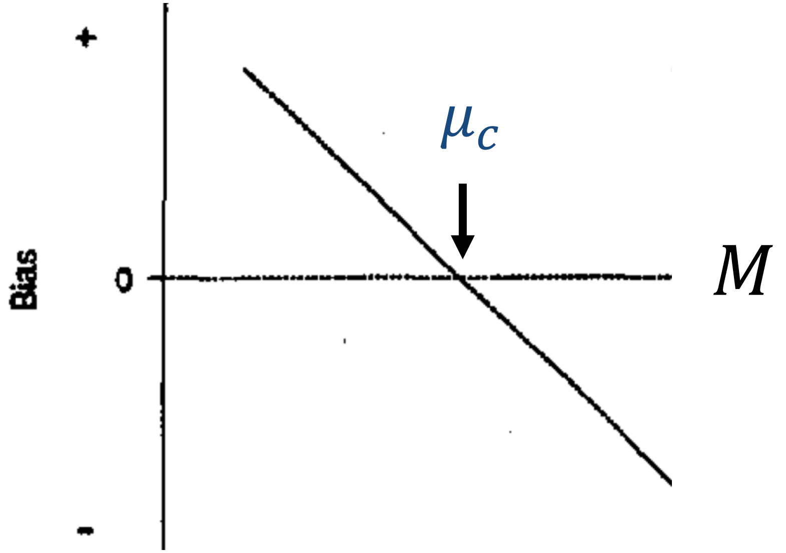
M is a noisy memory centered around true stimulus T. $M|T \sim N(T, \sigma_T^2)$

Reconstruction of the true stimulus T. $\hat{T} = w\mu_c + (1-w)M$

$$w = \frac{\sigma_T^2}{\sigma_c^2 + \sigma_T^2}$$



Category Adjustment Model (CAM) (Huttenlocher et al., 2000)



$$\hat{T} = w\mu_c + (1 - w)M$$

$\hat{T}$: Human Recall, μ_c: Category Center, M: Noisy Memory, w: Constant

Recall the size of a very **large** apple

→ 4 inches?

People's estimation shows little bias.

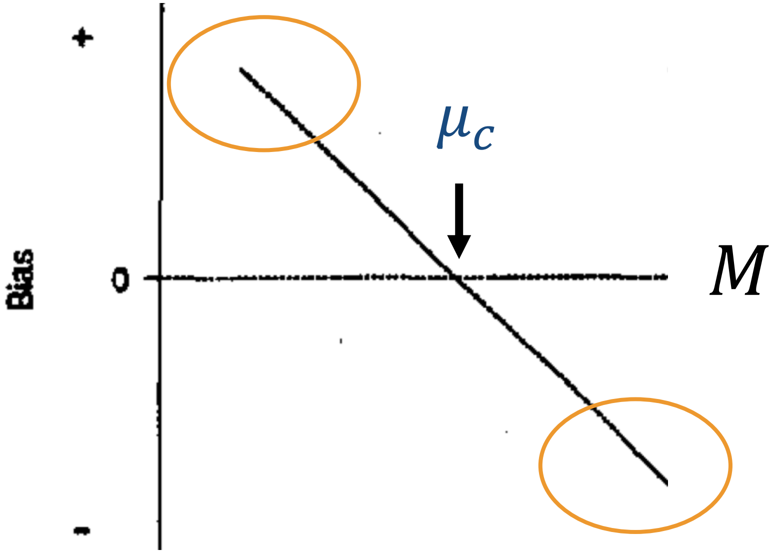
Atypical Item Effect

4 inches?

Atypical items will be less biased towards category center, compared with typical items.

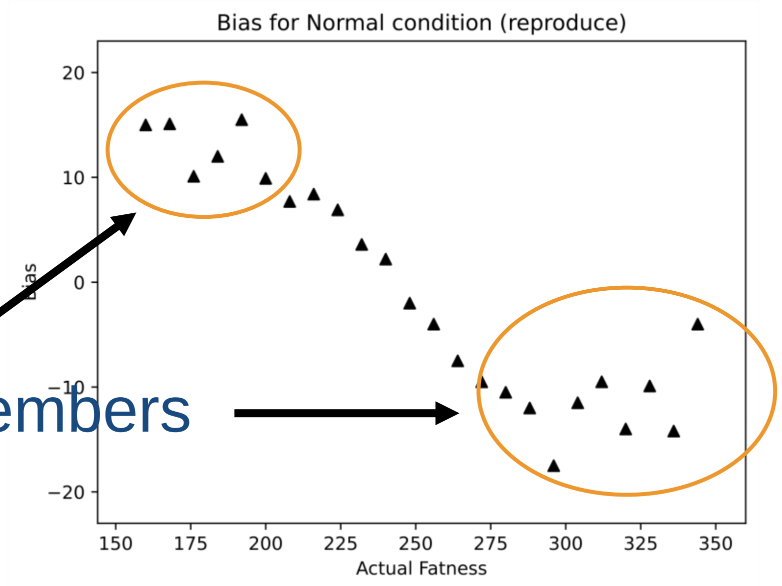
CAM model vs Atypical Item Effect

CAM model prediction



$$\hat{T} = w\mu_c + (1-w)M$$

Real world experiment



Atypical members

Huttenlocher et al., 2000

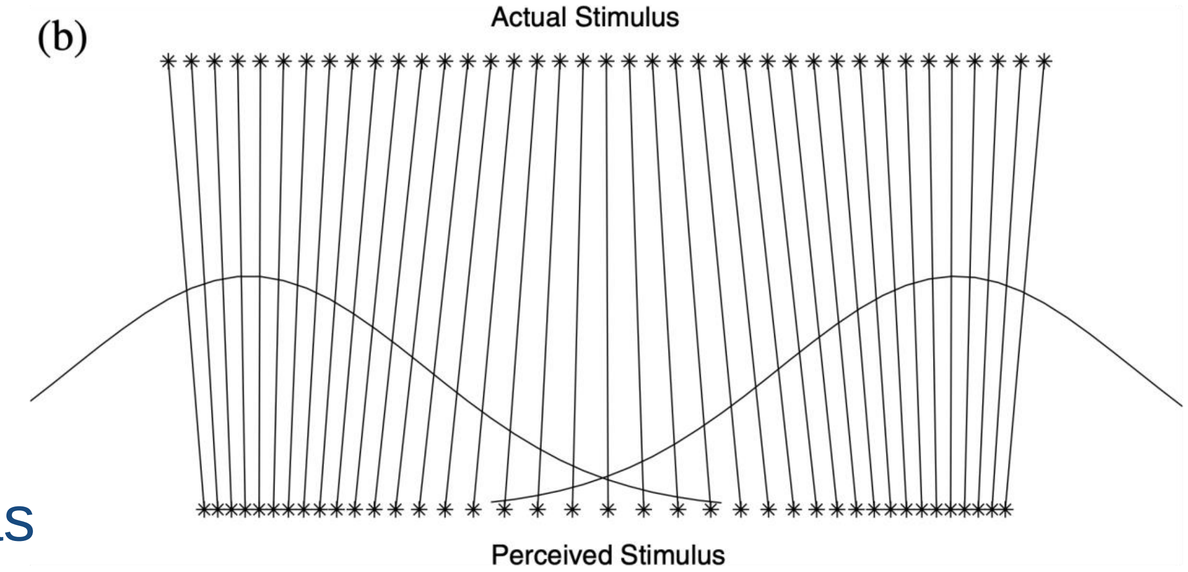
The CAM model cannot explain the Atypical Item Effect!

Other issues: multiple models for different Category Effects

- Multiple categories co-exist: Perceptual Magnet Effect
- Categories have multiple levels of abstraction: Hierarchical Category Effect

Perceptual Magnet Effect

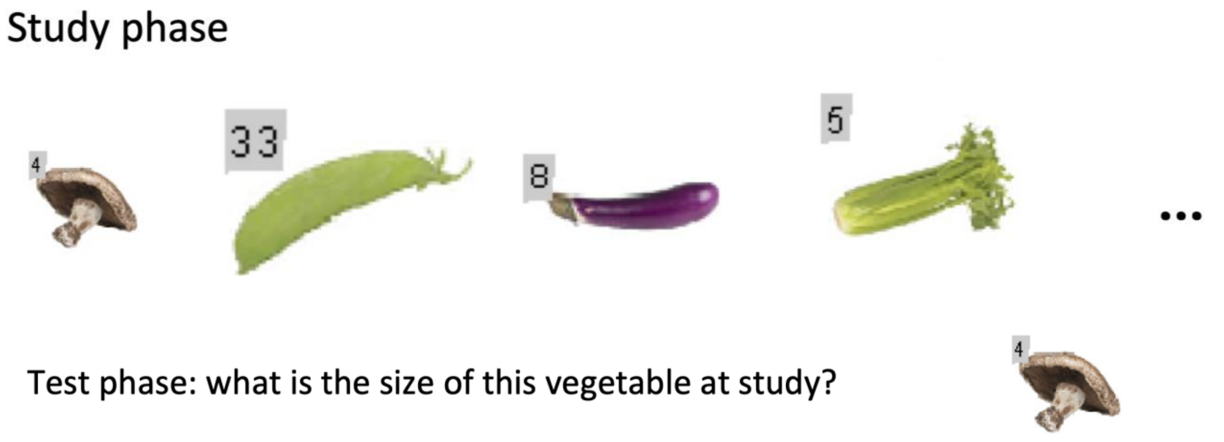
- Easier to discriminate speech sounds near a nonprototype of a category than near the prototype (Kuhl, 1991).
- It is as if there is shrinking of perceptual space near category prototypes, with prototypes acting as perceptual magnets that pull in the neighboring stimuli (Kuhl, 1993).



Feldman and Griffiths, 2007

Hierarchical Category Effect

- One can draw from prior knowledge of either mushroom size (object) or vegetable size (category)?
- Depending on the familiarity of the object!



Hemmer and Steyvers, 2009

Problems about current models on Category Effect

- Cannot explain the Atypical Item Effect
- Multiple models for different Category Effects

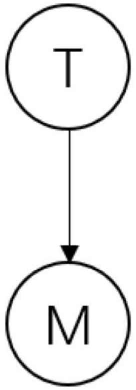
~~A set of Bayesian models that are each developed for a specific task.~~

A single Bayesian model that can explain a range of experimental effects!

Generalized-CAM model

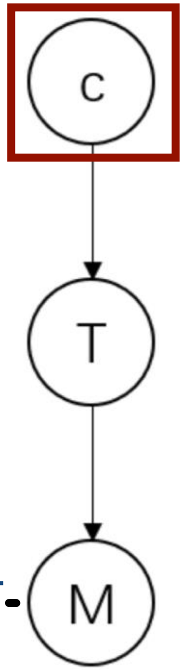
$$T \sim N(\mu_c, \sigma_c^2)$$

$$M|T \sim N(T, \sigma_T^2)$$



T is drawn from the category c

M is a noisy memory centered around T



$$T \sim N(\mu_c, \sigma_c^2)$$

$$M|T \sim N(T, \sigma_T^2)$$

$$E[T|M] = w\mu_c + (1-w)M$$

Reconstruction of T

$$E[T|M] = \Sigma_c E[T|c, M] * p(c|M)$$

Need to infer category c!

Generalized-CAM model

Calculation of $E[T|c, M]$

To calculate $E[T|c, M]$, we will prove that: $p(T|c, M)$ is a normal distribution, with a mean of $\frac{(\mu_c\sigma_T^2+\sigma_c^2 M)}{\sigma_T^2+\sigma_c^2}$.

Notice that the generative process in Equation 1 and 2 is a typical normal-normal model, we can obtain the distribution analytically with the following deduction.

Firstly, we will decompose $p(T|c, M)$ using Bayes rule. Since we assume that $p(c)$ has a uniform prior distribution and the likelihood $p(c, M)$ does not change with respect to the variable T, the posterior distribution is proportional to $p(M|T)p(T|c)$.

$$p(T|M, c) = \frac{p(M|T)p(T|c)p(c)}{p(M, c)} \propto p(M|T)p(T|c) \tag{A.5}$$

Now we further plug in the parameter $\sigma_c, \mu_c, \sigma_T$ in Equation 1 and 2 into the Normal Distribution formula and we obtain the expression of the posterior distribution as shown in Equation A.6.

$$\begin{aligned} p(T|c) &= \frac{1}{\sqrt{2\pi\sigma_c^2}} \exp\left\{-\frac{(T-\mu_c)^2}{2\sigma_c^2}\right\} \\ p(M|T) &= \frac{1}{\sqrt{2\pi\sigma_T^2}} \exp\left\{-\frac{(M-T)^2}{2\sigma_T^2}\right\} \\ p(T|M,c) &\propto p(M|T)p(T|c) \\ &= \frac{1}{\sqrt{\sigma_c^2\sigma_T^2}} \exp\left\{\frac{-(T-\mu_c)^2}{2\sigma_c^2} + \frac{-(T-M)^2}{2\sigma_T^2}\right\} \end{aligned} \tag{A.6}$$

We can then convert the term inside the exponential function to a quadratic form shown in Equation A.7:

$$\begin{aligned} &= \exp\left\{-\frac{1}{2}\left(\frac{T^2-2\mu_c T+\mu_c^2}{\sigma_c^2} + \frac{M^2-2MT+T^2}{\sigma_T^2}\right)\right\} \\ &= \exp\left\{-\frac{1}{2}\frac{\sigma_T^2 T^2 - 2\mu_c\sigma_T^2 T + \mu_c^2\sigma_T^2 + \sigma_c^2 M^2 - 2\sigma_c^2 MT + \sigma_c^2 T^2}{\sigma_c^2+\sigma_T^2}\right\} \\ &\propto \exp\left\{\frac{(\sigma_T^2+\sigma_c^2)T^2 - 2(\mu_c\sigma_T^2+\sigma_c^2 M)T}{\sigma_c^2+\sigma_T^2}\right\} \\ &\propto \exp\left\{\frac{\left(T - \frac{(\mu_c\sigma_T^2+\sigma_c^2 M)}{\sigma_T^2+\sigma_c^2}\right)^2}{\frac{\sigma_c^2\sigma_T^2}{\sigma_M^2+\sigma_c^2}}\right\} \end{aligned} \tag{A.7}$$

This concludes the proof that the distribution of $p(T|c, M)$ is a normal distribution, with mean $\frac{\mu_c\sigma_T^2+\sigma_c^2 M}{\sigma_T^2+\sigma_c^2}$.

Calculation of $p(c|M)$

$$\begin{aligned} p(M,c) &= \int_T p(c,M,T)dT \\ &= \int_T P(T|c)p(M|T)dT \\ &= \int_{-\infty}^{+\infty} \frac{1}{2\pi\sigma_T\sigma_c} \exp\left(-\frac{(T-\mu_c)^2}{2\sigma_c^2} - \frac{(M-T)^2}{2\sigma_T^2}\right) dT \end{aligned} \tag{A.1}$$

Given the fact that the following formula holds for $\forall a \in \mathbb{R}^+, b \in \mathbb{R}, C \in \mathbb{R}$ [1]:

$$\int_{-\infty}^{+\infty} e^{-ax^2-bx-m}dx = \frac{\sqrt{\pi}e^{\frac{b^2-4am}{4a}}}{\sqrt{a}} \quad (a > 0) \tag{A.2}$$

From equation A.1, we can derive that:

$$a = \frac{1}{2\sigma_c^2} + \frac{1}{2\sigma_T^2}, b = -\frac{\mu_c}{\sigma_c^2} - \frac{M}{\sigma_T^2}, m = \frac{\mu_c^2}{2\sigma_c^2} + \frac{M^2}{2\sigma_T^2} \tag{A.3}$$

Combining equation A.2 and A.3, we can derive $p(M, c)$ and $p(c|M)$.

$$\begin{aligned} p(M,c) &= \int_{-\infty}^{+\infty} \frac{1}{2\pi\sigma_T\sigma_c} \exp\left(-\frac{(T-\mu_c)^2}{2\sigma_c^2} - \frac{(M-T)^2}{2\sigma_T^2}\right) dT \\ &= \frac{1}{2\pi\sigma_T\sigma_c} \frac{\sqrt{\pi}}{\sqrt{\frac{1}{2\sigma_T^2}+\frac{1}{2\sigma_c^2}}} \exp\left(-\frac{(\mu_c-M)^2}{2\sigma_T^2+2\sigma_c^2}\right) \\ &= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_T^2+\sigma_c^2}} \exp\left(-\frac{(\mu_c-M)^2}{2\sigma_T^2+2\sigma_c^2}\right) \end{aligned} \tag{A.4}$$

$$p(c|M) \propto N\left(\mu_c, \sigma_T^2+\sigma_c^2\right)$$

$$\begin{aligned} E[T|M] &= \Sigma_c E[T|c, M] * p(c|M) \\ &= \Sigma_c \frac{\sigma_T^2\mu_c + \sigma_c^2 M}{\sigma_T^2+\sigma_c^2} * p(c|M) \end{aligned}$$

Key Insights

$$E[T|M] = \Sigma_c E[T|c, M] * p(c|M)$$

$$= \Sigma_c \frac{\sigma_T^2 \boxed{\mu_c} + \sigma_c^2 \boxed{M}}{\sigma_T^2 + \sigma_c^2} * \boxed{p(c|M)}$$

- The reconstruction of stimulus is first weighted by the inferred category membership $p(c|M)$, out of any number of possible categories.
- Under that category membership c, reconstruction then becomes a weighted combination between the category mean μ_c and the noisy memory M.

Three Category Effects

Category Effects we will cover

- Atypical Item Effect
- Perceptual Magnet Effect
- Hierarchical Category Effect

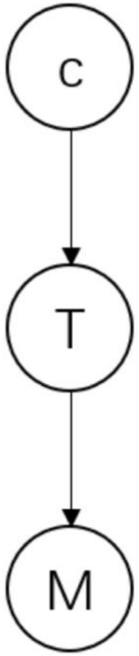
Intuition

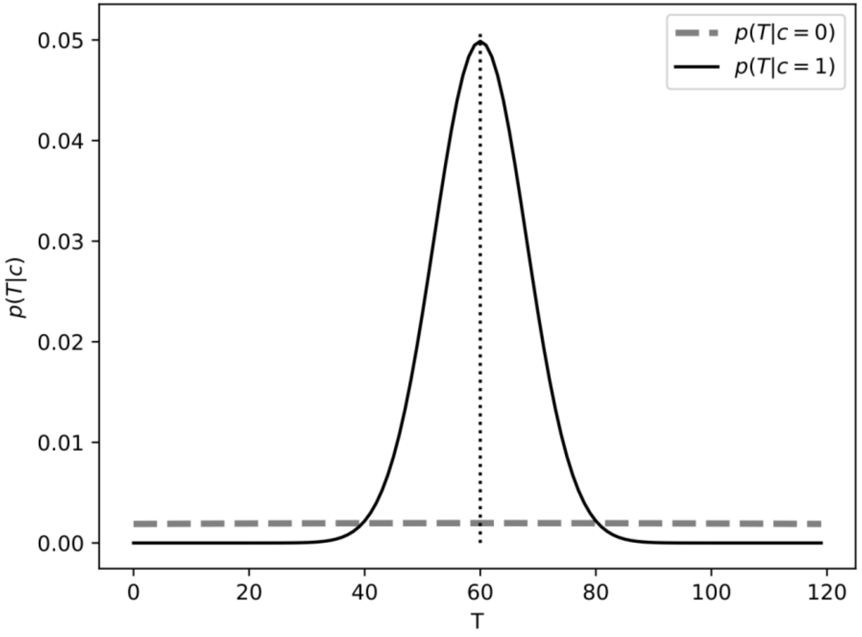
4 inches?

People actively infer whether a stimulus belongs to that category ($c = 1$) or not belong to that category ($c = 0$).

Atypical Item Effect

Implementation



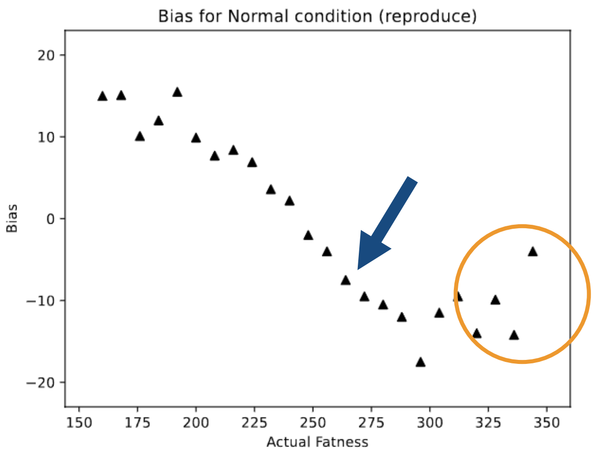


$$\sigma_{c=0} \gg \sigma_{c=1}$$

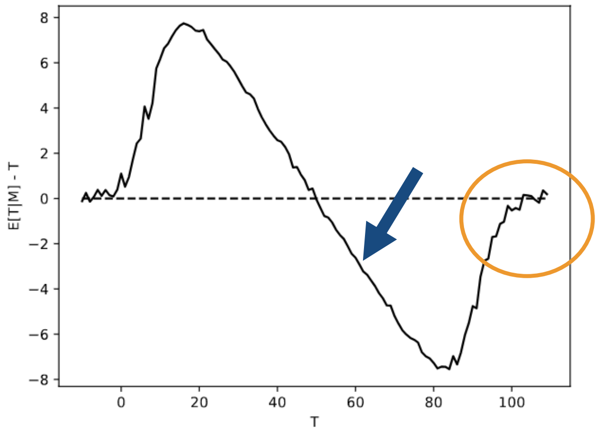
- Set one category as typical examples (c = 1), and a second "category" as atypical examples (c = 0), with large variance.
- Typical Items draw from $p(T|c=1)$.
- ATypical Items draw from $p(T|c=0)$.

Simulation Result

Experiment
(Huttenlocher et al., 2000)



G-CAM



- If inferred as typical, biased towards the category mean.
- If inferred as atypical, there is less bias.

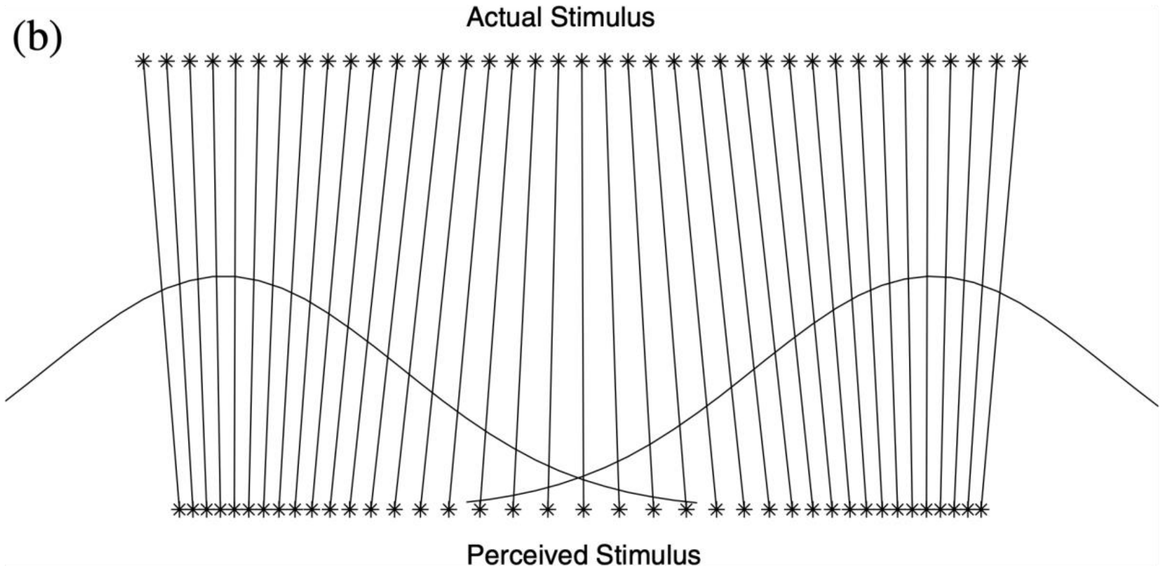
Reconstruction of T:

$$E[T|M] = p(c=1|M) * (w\mu_c + (1-w)M) + p(c=0|M) * M$$

Same stimulus reconstruction term in CAM

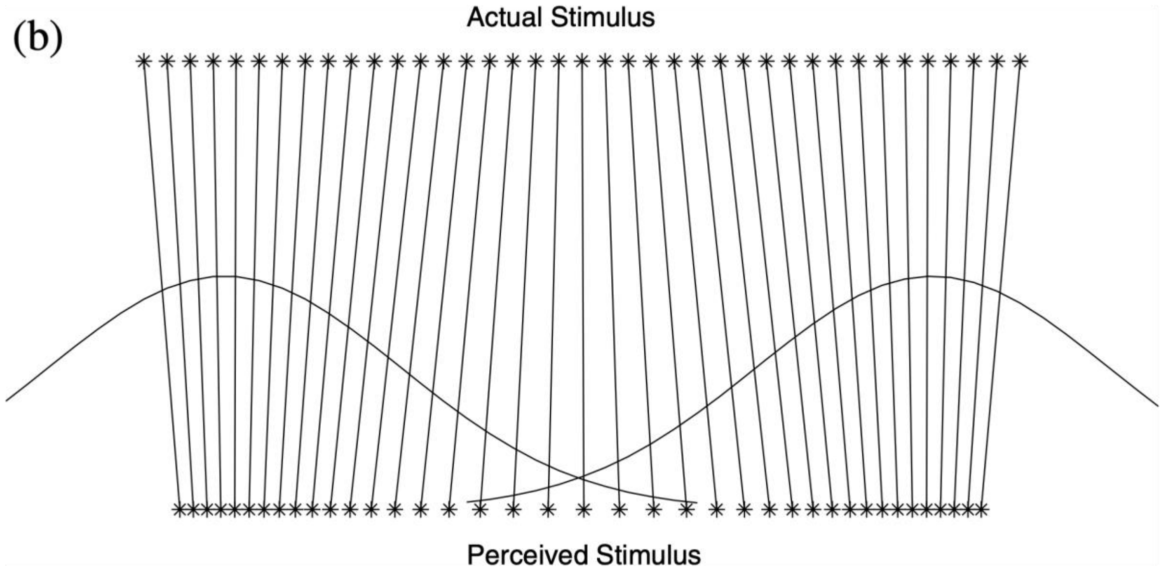
Intuition

Two or multiple-category case for the G-CAM model.

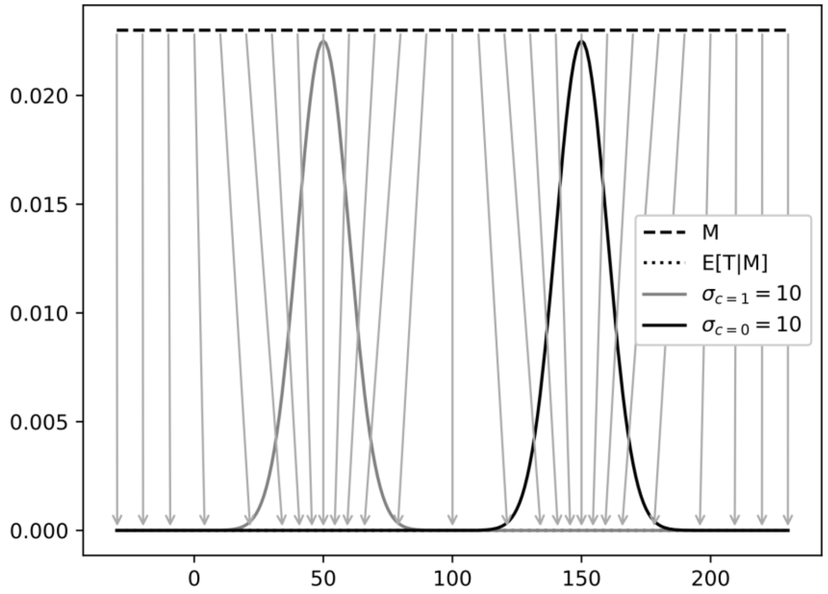


Feldman and Griffiths, 2007

Simulation



Feldman and Griffiths, 2007



Original Stimulus

E[T|M]: Perceived stimulus

Intuition

First infer the category of the stimulus, then reconstruct the stimulus based on the chosen category prior.

Object-level / category-level prior

A Bayesian model of reconstructive memory

(Hemmer and Steyvers, 2009)

- A mixture model where the prior mean and variance is a combination of object-level priors and category-level priors, weighted by familiarity z.

- Different than CAM, it treats memory noise as an unobserved variable.

$$T \sim N(\mu_0, \sigma_0)$$

$$\mu_0 = z\mu_i + (1 - z)\mu_c$$

$$\sigma_0^2 = z\sigma_i^2 + (1 - z)\sigma_c^2$$

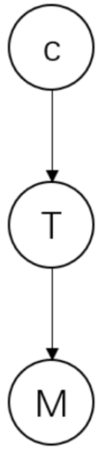
Hierarchical Category Effect

Our model

- When a stimuli is considered part of a object membership (o = 1), it is biased towards the object mean.

$$T|o=1 \sim N(\mu_o, \sigma_o^2) \qquad E[T \mid o=1, M] = \frac{M\sigma_o^2 + \mu_o \cdot \sigma_T^2}{\sigma_o^2 + \sigma_T^2}$$

Familiar



- When a stimuli is identified as not from the object membership (o = 0), it will resort to the prior information from the higher-level category mean.

$$T|o=0 \sim N(\mu_c, \sigma_c^2) \qquad E[T \mid o=0, M] = \frac{M\sigma_c^2 + \mu_c \cdot \sigma_T^2}{\sigma_c^2 + \sigma_T^2}$$

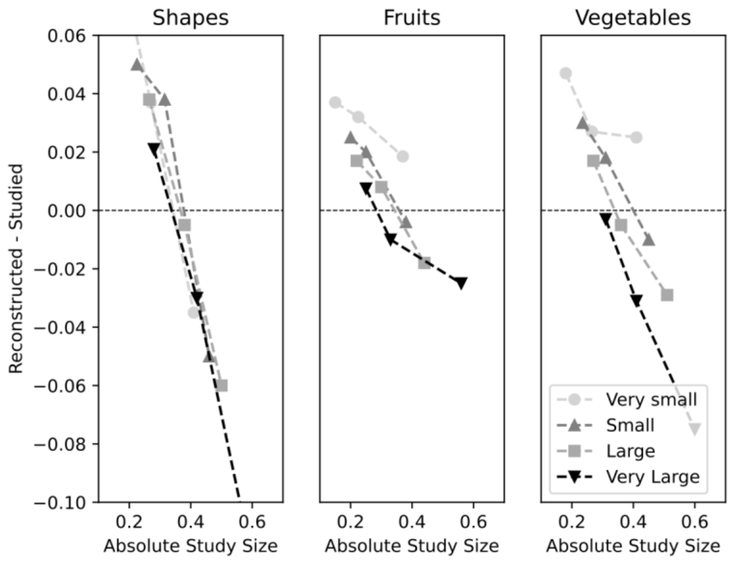
Not familiar

- Same memory noise term as before. $M|T \sim N(T, \sigma_T^2)$

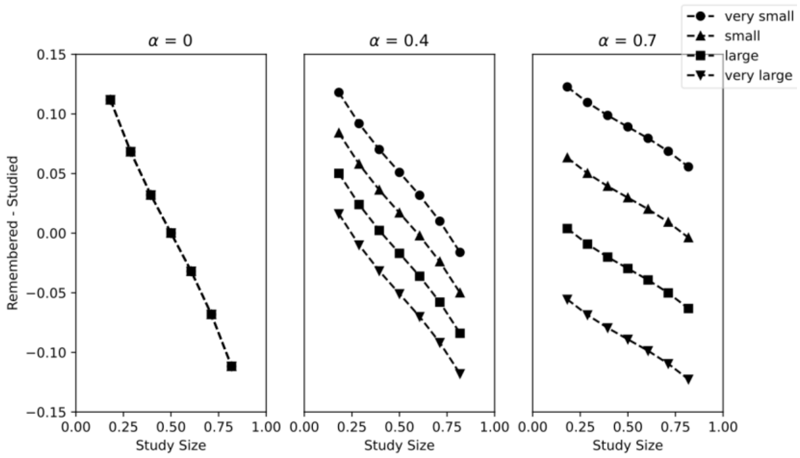
Simulation

Experiment

(Hemmer and Steyvers, 2009)



G-CAM



Summary

- We proposed a generalized CAM model.
- It can account for empirical findings of atypical examples while unifying previous models of category effects.

Acknowledgements

Qiong Zhang

Assistant Professor in Psychology & Computer Science

Pernille Hemmer

Director of the Human Computational Cognition Laboratory

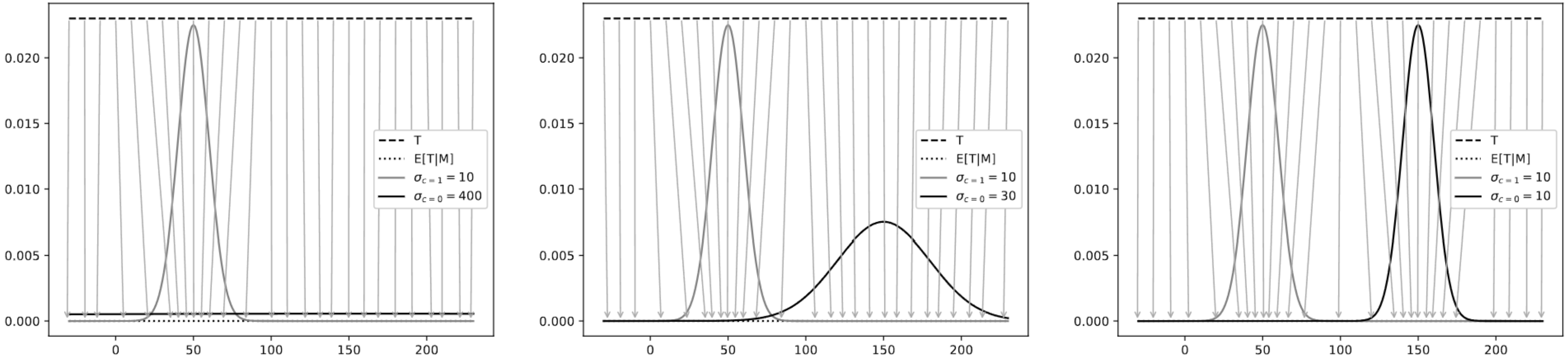
RUTGERS
THE STATE UNIVERSITY OF NEW JERSEY

> MEMORY OPTIMIZATION LAB

Thank you & Any questions?

zihao.xu@rutgers.edu
Paper: https://osf.io/preprints/psyarxiv/9a7ft/

Simulation: Atypical Item Effect and Perceptual Magnet Effect



Original Stimulus

E[T|M]: Perceived stimulus