



Oddness at a glance: Unraveling the time course of typical and atypical scene perception

Abraham Botros¹, Michelle R. Greene¹, & Li Fei-Fei¹

¹: Stanford University, Department of Computer Science



Background and Motivation

- We rapidly recognize scene gist [1-4], but the mechanism is unknown.
- How do we do it? To what extent does typicality and expectation aid perception?

Methods I – An Odd Database

- 100 images: 50 atypical ("odd") images, 50 visually-similar typical ("doppelgänger") images

Typical (Doppelgänger)



Atypical (Odd)



Methods II – Free Response System

- 10 participants, 50 images each
- Presentation times: 20ms, 40ms, 80ms, 150ms, 500ms
- Dynamic naturalistic masks: 80ms total
- Free responses [1]



PT: 20-500ms



Masks: 80ms total

Free response

Unrestricted free response

Free Response Example I



Typical

20ms: *Three people sitting and laughing in a cafe.*



Atypical

A boy and a girl sit facing one another at a table as the girl eats some ice cream. She is wearing a dress.

This shows two girls seated around a round white table in what may be a fast-food restaurant. They are asian. The girl on the left is wearing a black polka-dotted skirt with white dots. The skirt is to her knee. The girl on the right is wearing short. The scene is red color in the background. There is a low wall blocking the table and the girls from the rest of the restaurant.

500ms: *3 ladies at a table. they are all looking at one another. Looks like a balcony table. a bag on the floor. a carton of juice on the table*

Two asian girls seated facing one another. Girl on left reading book, girl on right looking at her phone

Free Response Example II



Typical

40ms: *Sheet rock, the type that often lies next to highways in the Allegheny area of Pennsylvania*

150ms: *This is a large grey jagged boulder on a grassy open space in a natural habitat with green trees and blue sky.*

500ms: *Large boulder next to mountain range*



Atypical

Some people standing in a living room looking out a screen door.

a man standing next to some sort of stone face or statue inside a building

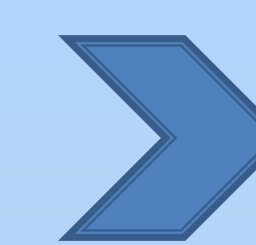
3 people in a room. maybe a living room where i guess a large stone fell down through the roof and landed in the living room right next to a couch. The floor is split in half

Methods III – Crowd-Sourced Analysis

Objects, Scenes



Attributes, Descriptors



Correctness Analysis

Completeness Analysis

Amazon Mechanical Turk (AMT) - large-scale, crowd-sourced labeling and analysis of free responses

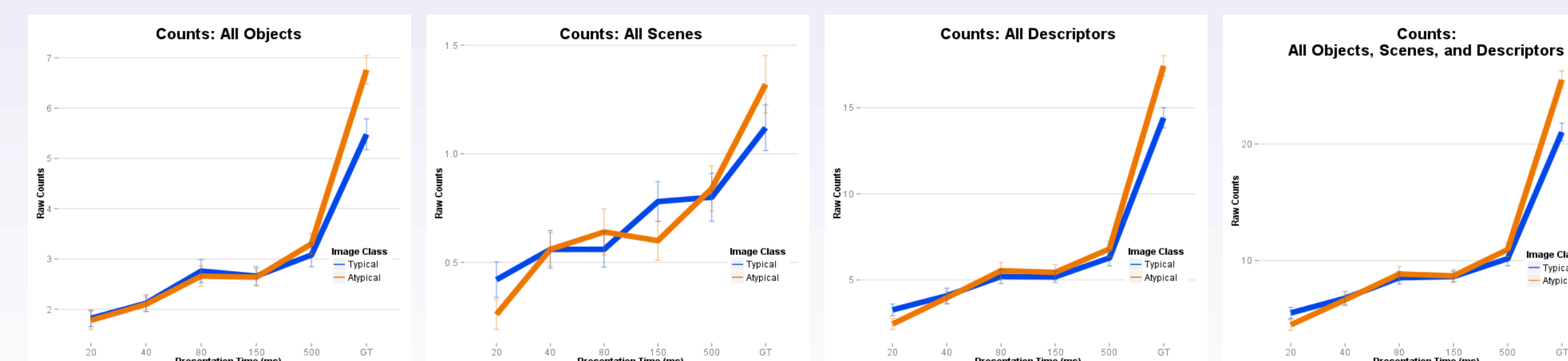
Results and Observations

Atypicality Decreases Overall Description Rating



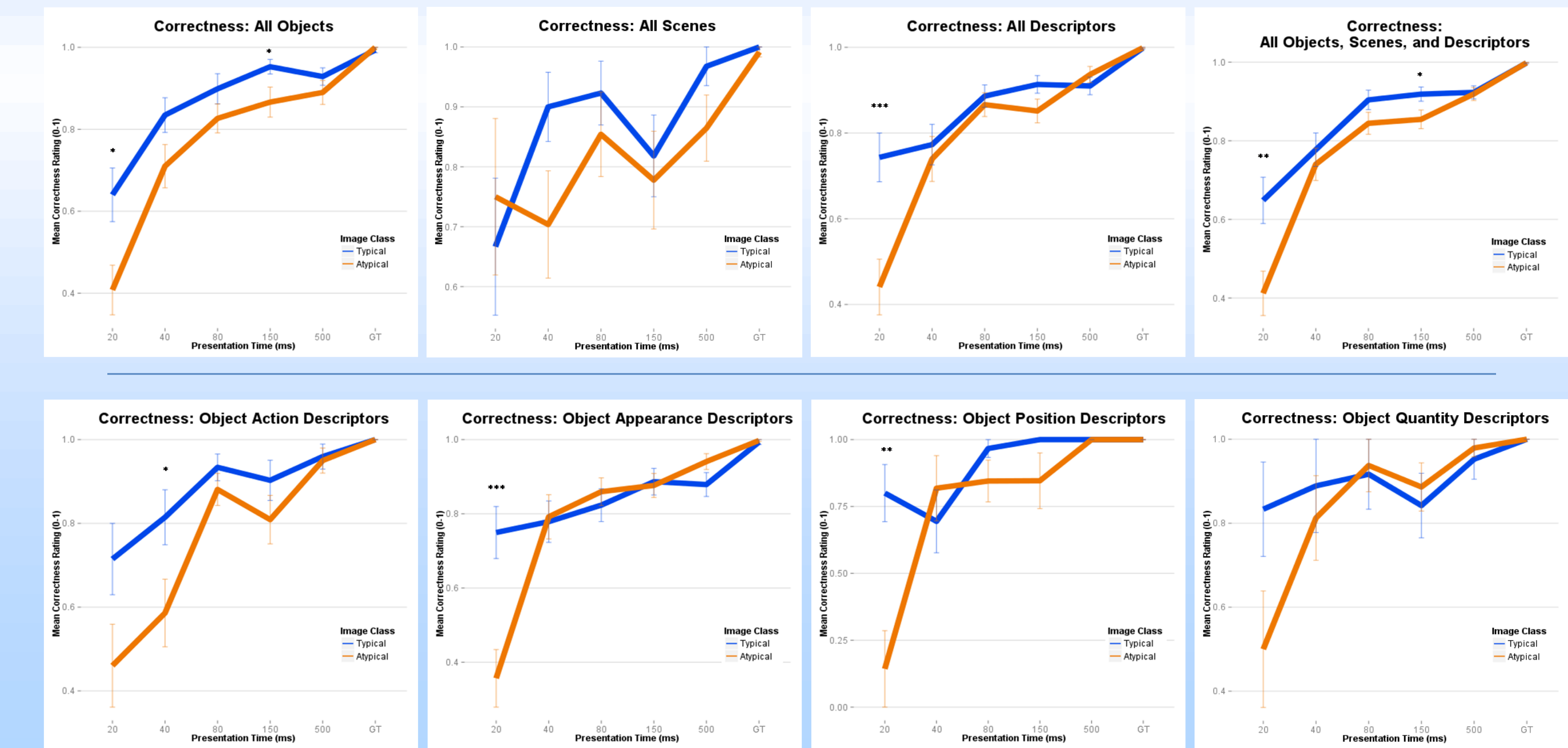
AMT workers rated overall correctness and thoroughness according to given guidelines ($F(1,598)=7.3$, $p<0.01$). In addition, for atypical images, they rated how well the writer perceived and understood the oddity.

Similar Key-Term Counts for Typical and Atypical Images



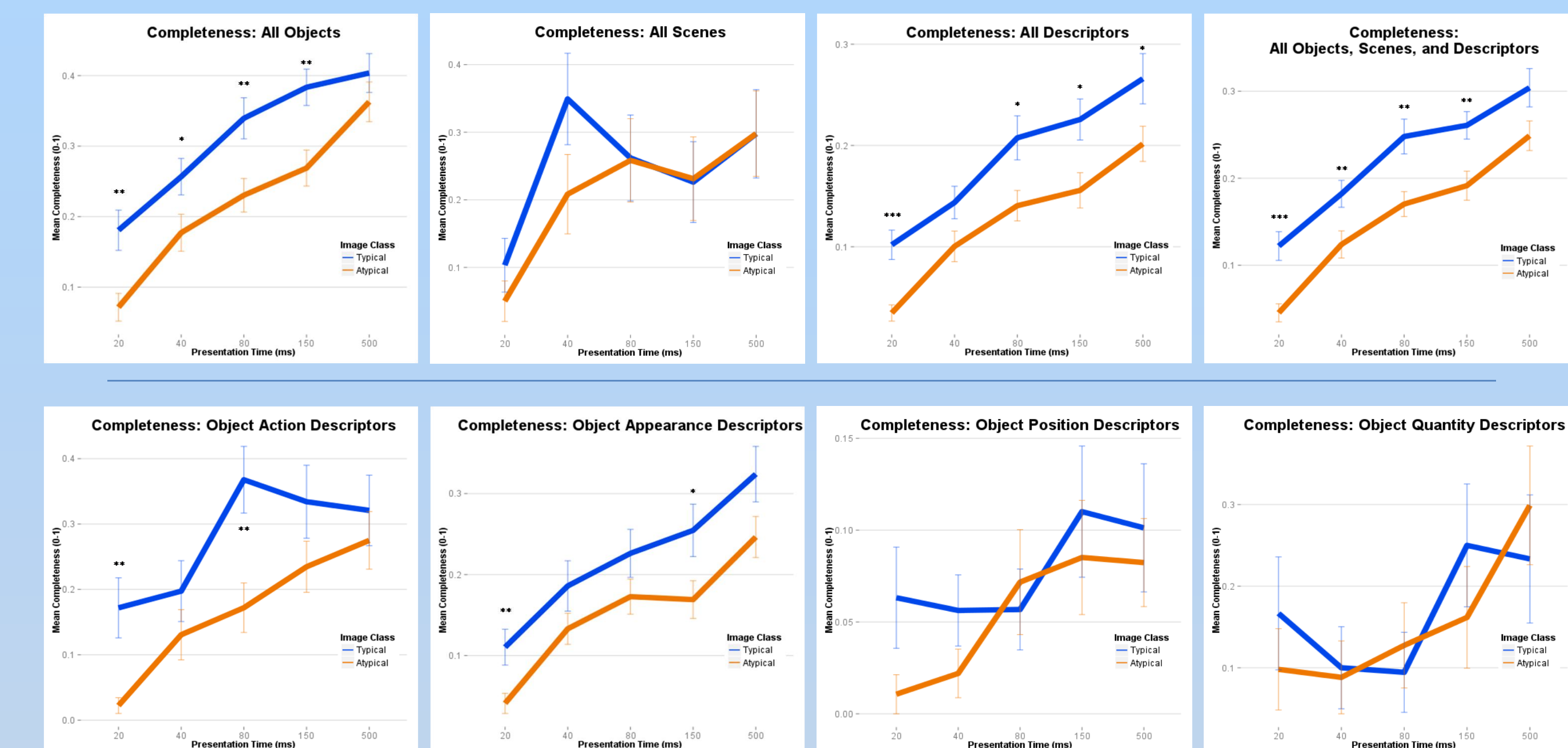
Raw counts of the number of key terms extracted by AMT workers from descriptions. No significant differences between typical and atypical responses.

Atypicality Decreases Individual Key-Term Correctness



Top row: Average correctness ratings for each specified category. Reported details were significantly worse for odd images, especially at earlier presentation times. Bottom row: Average correctness ratings for object descriptors only, divided into the four specified descriptor categories (action, appearance, position, quantity).

Atypicality Decreases Description Completeness



Top row: In addition to rating the correctness of what observers wrote, AMT workers also analyzed what aspects of the images they failed to describe. Compared to typical images, observers viewing odd images missed significantly more details at nearly every presentation time.

Conclusions and Discussion

- Results suggest a significant role of typicality and expectation in early visual processing.
- Despite similar key term counts, typical and atypical descriptions differed significantly in several critical facets.
- Rationalizing and neglect of oddities at shorter PTs (20ms-150ms)
- Atypical image descriptions suffered correctness and completeness penalties at shorter PTs, but recovered by 150ms.
- Substantial delay for atypical images. Why is odd stimulus processing delayed?
 - Need for additional early processing and feedback?
- Future work includes determining the extent to which these effects are perceptual versus conceptual.

References:
1. Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision*, 7(1:10), 1-29.
2. Greene, M. R., & Oliva, A. (2009). The Briefest of Glances: The Time Course of Natural Scene Understanding. *Psychological Science*, 20, 464-472.
3. Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, 2(5), 509-522.
4. Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381, 520-522.