Modeling the Influence of Situational Variation on Theory of Mind

Wilka Carvalho
Mentors: Ralph Adolphs, Bob Spunt, and Damian Stanley

Abstract

Recent advances in neuroimaging techniques and decision-making research are paving the way for research on the brain and insight into its function. This study is part of an ongoing project to use model-based behavioral and functional Magnetic Resonance Imaging (fMRI) techniques to understand the neurocognitive processes that allow people to make inferences about the minds of other people. We focus on how people use mental representations of other people supplemented by trait inferences to predict their behavior. We developed the Trait Learning Task (TLT), in which participants (Judges) learn about the distinguishing traits of other people (Actors) by observing how each person responds behaviorally to various stimuli (Situations). This task enables analysis of the process by which a Judge learns a trait. We are administering this test to neurotypical and non-neurotypical individuals such as those with Autism Spectrum Disorder (ASD). Task performance will be used to fit Bayesian mathematical models of the computations necessary for social learning. We hope to learn about how social norms, individual behavior, and mental representations are integrated to infer traits.

Introduction

Imagine you see a woman, Heather, with a gun being pointed at her. Heather's face expresses fear. It is expected that facing a gun would induce fear—thus, you find Heather's behavior normal. Now, in place of fear, imagine the gun instead induced laughter. This expression counters your expectation, so you take a mental note, "Heather did not respond as expected to the situation," and perhaps infer some trait about her—that she is skeptical, optimistic, or just strange.

In this process, you have used a cognitive ability known as Theory of Mind, and information on the norm behavioral response to the situation, to infer a trait of Heather's by means of a cognitive process known as Attribution. Theory of Mind (ToM) is the process of creating or maintaining a mental representation of another person, which can be used to create mental state inferences for that person (Malle, 2008). Attribution is the process of integrating situational and behavioral information to infer the traits of another (Barnhill, 2006; Gilbert et al, 1995; Kelly, 1973). Together, the two facilitate coordinated social behavior—fundamental to building relationships and maintaining mental well-being, exemplified by non-neurotypical individuals with Autism Spectrum Disorder (ASD) that have shown impairment in ToM, linked to social dysfunction—often their most significant complaint in everyday life (Kennedy et al, 2012).

This project begins a larger project aimed at studying the influence of situational norms on Attribution and ToM. We focus on how ToM and Attribution are used to infer traits related to stress, specifically, such fear and sadness. We developed a behavioral model describing the role of Attribution in trait learning, and an experimental paradigm, coined the "Trait Learning Test," (TLT), to test it. Currently, no free, web-based general-purpose platforms able to administer experiments with user-input contingent progression exist. We developed a web application, coined the, "New Experimental Tool for

Psychology (NEXT Psych) Web Application," designed for experiments that require user-input contingent screen and trial progression, and data presentation.

We will create a computational model for cognitive processes of ToM and Attribution. Recent research has shown that (functional Magnetic Resonance Imaging) fMRI recorded data for neural activity of a cognitive process' neural region or network is consistent with that process' corresponding computational model (O'Doherty, 2011). A distinct network of interconnected brain regions known as the ToM network subserves ToM(Kennedy et al, 2012), indicating that fMRI scans may be used to understand the neural processes that occur in the brain as a person uses ToM and Attribution. We will apply model-based fMRI techniques towards the study of the neural computations necessary for trait attribution, for the construction/update of mental representations, and to identify the neural bases of these computations.

Behavioral Schematic Development

We define the person using ToM as the “Judge,” and the person on whom ToM is used as the “Actor”. We define “Situational” information as the contextual information presented along with an Actor, while “Behavioral” information is the information presented relevant to the Actor’s reaction to the Situation. “Situational” information encompasses the situation’s behavioral “norms”. Behavioral norms are the behavioral responses associated with, and considered “normal” for, a situation. Keeping to our example: You are the Judge, Heather is the Actor, ‘facing a gun’ is the situational information, responding with fear is the situational norm, Heather’s reaction of fear or laughter is the behavioral information. (Gilbert, 1995) Events of the Attribution process are detailed as:

a) Recognizing the situation,
b) Bringing in associated beliefs for the expected (norm) behavior in the situation,
c) Perceiving and categorizing the observed individual's behavior, and
d) Determining whether the Actor's behavior violates behavioral expectations.

As a Judge, noting that Heather is facing a gun would correspond to (a); that fear is a norm behavior, (b). Here you make a prediction on what Heather’s behavioral response will be. With no previous information on Heather, you create a behavioral expectation entirely from the situation’s behavioral norms, expecting Heather to respond with fear. Noting that Heather responded with laughter corresponds to (c); and that laughter is an unexpected behavior to (d). Heather’s behavior has violated your behavioral expectation, deviating largely from fear, the situation’s behavioral norm. The measurement of how much Heather’s (the Actor’s) behavior violates the

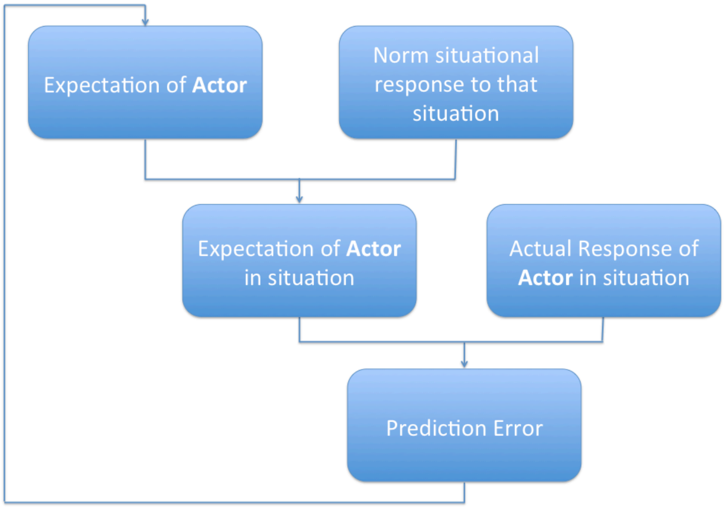

Expectation of Actor
Norm situational response to that situation
Expectation of Actor in situation
Actual Response of Actor in situation
Prediction Error

Figure 1: A structural model illustrating the behavioral paradigm for inferring a trait.

behavioral expectations is known as the prediction error. This is used to infer traits about Heather and update your (the Judge's) corresponding mental representation. The larger the prediction error, the more diagnostic it is of a possible trait, and the more consequential the update to a Judge's mental representation. The smaller the prediction error, the less diagnostic of a possible trait, and the less consequential the update. The next time that you see Heather in a situation with fear as a norm behavioral response, you less likely expect Heather to respond with fear, and more likely expect her to respond with a behavior similar to laughter. Continual exposure to an Actor in various situations allows a Judge to continue updating his/her mental representation for the Actor. In future predictions, the Judge incorporates the relevant inferred traits in their prediction of an Actor's behavior, which, presumably, allows for more accurate future behavioral predictions.

NEXT Psych Web Application Development

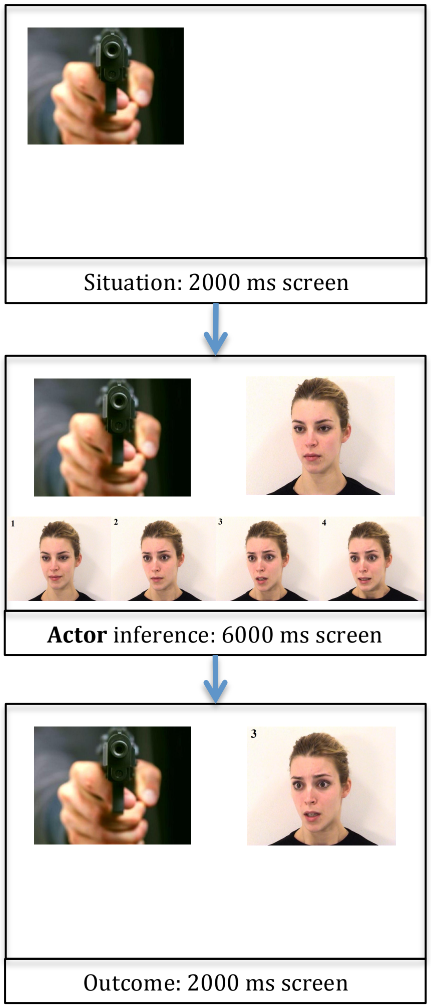
NEXT Psych was developed to support presentation of various text or media "objects" and progression dependent on time and key-input. It is designed to run what we refer to as "Blocks." Blocks are collections of "Trials." Trials are instances in which the user performs some task. A Trial consists of the "Events" required to perform the task. Events are divided into six sub-events types: "Clear," "Timed," "Key," "TimedKey," "TimedOrKey," and "Feedback." The "Clear" sub-event clears any chosen objects from the screen, and is intended to update aspects of the screen or to begin entirely new trials. All other sub-events allow the presentation of an object on the screen. Each sub-event has a specific event it waits for before it progresses to the next sub-event; "Timed" waits for a period of time to pass; "Key" until an accepted key is pressed; "TimedKey" a period of time for an accepted key to be pressed; "TimedOrKey" a period of time or until an accepted key is pressed. "Feedback" is a user-input contingent sub-event that performs precisely the same operation as another sub-event that it "mimics." To facilitate flexibility of this application, the module that runs the experiment has been created in such a way that it easy to add sub-events types. This functionality, along with the already-defined sub-events and the dynamic object placement capabilities, provides the user great flexibility in object presentation and trial progression.

Trait Learning Test

We sectioned TLT into trials in which the Judge has an opportunity to infer on a single trait of the Actor's. Trials consist of the following events:

(1) The situational information is presented and the Judge is asked to provide how he/she would typically react to the situation.
(2) A neutral expression of the Actor is provided for reference along with an extremity-based gradient of mental states for a specific type of reaction. Here, the Judge is asked to choose which extremity level he/she believes is the most probable outcome.
(3) The Judge is then shown the actual response of the Actor.

There are multiple trials for each Actor, and multiple Actors. The trials corresponding to each Actor are grouped together and randomized – forcing the judge to rely on his/her working memory. After a certain number of trial presentations, event (3) is no longer presented. We present the trials to Judges to test how they attribute traits related towards stress, specifically. Event (1) lasts for 2000 ms, event (2) for 6000 ms or until the Judge chooses on an extremity level, and event (3) for 2000 ms. The first part (with event (3)) is meant to allow the Judge to create an Actor representation, while the second part (without event (3)) is used to collect data on the functional form of the learning curve of the Judge. We are interested in observing how Judges re-evaluate their initial attribution once they begin to recognize a pattern for how the Actor tends to respond to different types of stimuli with varying associated norm behavioral responses.


Situation: 2000 ms screen
1
2
3
4
Actor inference: 6000 ms screen
3
Outcome: 2000 ms screen

Methods

NEXT Psych has been written in JavaScript and PHP. It is being administered to 50 native English speaking U.S. residents recruited online through the crowdsourcing service, Mechanical Turk. In addition to performing the TLT, these participants will also complete a battery of questionnaires that measure demographic and personality characteristics relevant to social function. If the TLT measures a key component of social intelligence, individual differences in TLT performance should be related to scores on these measures.

Situations for TLT were chosen from a stimulus set obtained from the Nencki Affective Picture System (NAPS) (Marchewka , Żurawski , Jednoróg , & Grabowska , 2013) and the emotional expressions of ten female Actors were chosen from a stimulus set obtained from the Amsterdam Dynamic Facial Expression Set (ADFES) (PRI, 2013). Situations were sorted by valence. Valence, here, refers to the intrinsic attractiveness (low valence) or aversiveness (high valence) of the Situations. Situations with medium or extreme valence values were eliminated to avoid data that may offset the results. The Situation set was created such that it contains a range in extremity levels of the associated norms. Another member of the group, Michael Belcher, has collected normative ratings on the stimulus sets in relation to stress levels related to fear and sadness.

Three Actors, 34 Situations corresponding to sadness, and 33 Situations corresponding to fear were chosen. For each Actor, we have a 4-point equidistance behavioral scale for the level of fear or sadness the Actor is exhibiting. For every situation, the most consistent chosen behavioral level in the scale is defined as the norm behavioral response. For every Actor, for every situation, the point in the scale most consistently chosen as the expected response is chosen as the actual response of the actor.

In a trial a Judge is asked to choose a point on the scale, P, that they expect to be the Actor's response. The difference between P and the Actor's actual response, A, is the Prediction Error, E so E = P-A In constructing the computational model for ToM and Attribution, we will find the factor by which E is used to update a Judge's behavioral expectation of an Actor for both fear-inducing and sadness-inducing situations.

On Caltech's Emotion and Social Cognition Laboratory and Greater

The intended implications of this project are two-fold. First, to advance the Laboratory's general efforts in understanding the neural basis of the psychological processes that allow us to make inferences about what happens inside the minds of other people (Adolphs, 2009). Second, to facilitate sophisticated online behavioral testing, especially in fields such as psychology and neuroscience. NEXT Psych has been written to support this experiment, and other current and future experiments of the laboratory. It is open-source, publicly available[1], and under active development—meant to benefit the larger scientific community.

Acknowledgments

I thank the Principle Investigator, Ralph Adolphs, for his help and guidance this summer. I thank the two post-doctoral scholars Bob Spunt and Damian Stanley for helping me familiarize myself with the topics and analytical approaches necessary for this work. I thank the Student-Faculty Programs of Caltech, specifically the Minority Undergraduate Research Fellowship (MURF), for this summer research opportunity. Lastly, I thank the Howard Hughes Medical Institute for funding this research.

References

Adolphs, R. (2009). The Social Brain: Neural Basis of Social Knowledge, 60(1), 693–716. doi:10.1146/annurev.psych.60.110707.163514

Barnhill, R. (2006). Attributions as Behaviour Explanations: Towards a New Theory, 1–24.

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*.

Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist*, *28*(2), 107–128. doi:10.1037/h0034225

Kennedy, D. P., & Adolphs, R. (2012). The social brain in psychiatric and neurological disorders. Trends in Cognitive Sciences, 16(11), 559–572. doi:10.1016/j.tics.2012.09.006

Malle, B. F. (2008). The Fundamental Tools, and Possibly Universals, of Human Social Cognition. In *Handbook of Motivation and Cognition Across Cultures* (pp. 267–296). Elsevier. doi:10.1016/B978-0-12-373694-9.00012-X

Marchewka , A., Żurawski , Ł., Jednoróg , K., & Grabowska , A. (2013). The Nencki Affective Picture System (NAPS): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database . Behavioral Response .

Moran, J. M., Young, L. L., Saxe, R., & Lee, S. M. (2011). Impaired theory of mind for moral judgment in high-functioning autism. Presented at the Proceedings of the …. doi:10.1073/pnas.1011734108/-/DCSupplemental

[1] https://github.com/wcarvalho/NEXT-Psych

O'Doherty, J. P., Hampton, A., & Kim, H. (2007). Model-Based fMRI and Its Application to Reward Learning and Decision Making. Annals of the New York Academy of Sciences, 1104(1), 35–53. doi:10.1196/annals.1390.022

Marchewka , A., Żurawski , Ł., Jednoróg , K., & Grabowska , A. (2013). The Nencki Affective Picture System (NAPS): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database . Behavioral Response .

Yoshida, W., Dziobek, I., Kliemann, D., Heekeren, H. R., Friston, K. J., & Dolan, R. J. (2010). Cooperation and Heterogeneity of the Autistic Mind. Journal of Neuroscience, 30(26), 8815–8818. doi:10.1523/JNEUROSCI.0400-10.2010