1. **REPORT DATE**  MAR 2015  
2. **REPORT TYPE**  
3. **DATES COVERED**  00-00-2015 to 00-00-2015  

4. **TITLE AND SUBTITLE**  
   **Understanding Second-Order Theory of Mind**  

5a. **CONTRACT NUMBER**  
5b. **GRANT NUMBER**  
5c. **PROGRAM ELEMENT NUMBER**  
5d. **PROJECT NUMBER**  
5e. **TASK NUMBER**  
5f. **WORK UNIT NUMBER**  

6. **AUTHOR(S)**  

7. **PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**  
   **Naval Research Laboratory**, 4555 Overlook Ave, SW, Washington, DC, 20375  

8. **PERFORMING ORGANIZATION REPORT NUMBER**  

9. **SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**  

10. **SPONSOR/MONITOR’S ACRONYM(S)**  
11. **SPONSOR/MONITOR’S REPORT NUMBER(S)**  

12. **DISTRIBUTION/AVAILABILITY STATEMENT**  
    Approved for public release; distribution unlimited  

13. **SUPPLEMENTARY NOTES**  

14. **ABSTRACT**  
    Theory of mind is a key factor in the effectiveness of robots and humans working together as a team. Here, we further our understanding of theory of mind by extending a theory of mind model to account for a more complicated, second-order theory of mind task. Ultimately, this will provide robots with a deeper understanding of their human teammates, enabling them to better perform in human-robot teams.  

15. **SUBJECT TERMS**  

16. **SECURITY CLASSIFICATION OF:**  
   a. **REPORT**  unclassified  
   b. **ABSTRACT**  unclassified  
   c. **THIS PAGE**  unclassified  

17. **LIMITATION OF ABSTRACT**  
   Same as Report (SAR)  

18. **NUMBER OF PAGES**  2  

19a. **NAME OF RESPONSIBLE PERSON**  

---

Reports published on or after March 1, 1995, must be handled in accordance with OMB Control Number 0704-0188. Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
Understanding Second-Order Theory of Mind

Laura M. Hiatt  
Naval Research Laboratory  
4555 Overlook Ave, SW  
Washington, DC 20375  
laura.hiatt@nrl.navy.mil

J. Gregory Trafton  
Naval Research Laboratory  
4555 Overlook Ave, SW  
Washington, DC 20375  
greg.trafton@nrl.navy.mil

ABSTRACT
Theory of mind is a key factor in the effectiveness of robots and humans working together as a team. Here, we further our understanding of theory of mind by extending a theory of mind model to account for a more complicated, second-order theory of mind task. Ultimately, this will provide robots with a deeper understanding of their human teammates, enabling them to better perform in human-robot teams.

Categories and Subject Descriptors
I.2.0 [Artificial Intelligence]: General—cognitive simulation; I.2.11 [Artificial Intelligence]: Distributed Artificial Intelligence—intelligent agents, coherence and coordination

General Terms
Theory

Keywords
theory of mind; human-robot teams

1. INTRODUCTION
Theory of mind (ToM) is a critical capability for teams of agents working together, whether robots, humans, or both; without it, people can be extremely limited in their abilities to interact naturally and effectively with others [3, 7]. By better understanding how people achieve this core component of teamwork, we can develop robots that are able to more effectively coordinate with their human partners, both by enabling them to have theory of mind in the same way, and, just as importantly, by allowing them to understand what beliefs, desires and intentions a human partner may be ascribing to them or to someone else [4].

Since adult humans are typically very adept at theory of mind tasks, studies on theory of mind’s underlying mechanisms are often performed on developing children. Additionally, studies are typically done on variations of false-belief (FB) tasks [7], a common setting in which to explore ToM where participants distinguish between a “true-belief” answer (e.g., where the participant thinks an object actually is) and a “false-belief” answer (e.g., where the participant believes someone else thinks an object is). More complicated variations include avoidance tasks, or predicting someone’s behavior based on their identified false-belief [6].

In previous work, we developed a cognitively-plausible process model of how children develop the ability to correctly answer false-belief and avoidance queries [5]. The model’s theory states that children simultaneously develop the abilities to answer these queries, including using cognitive simulation for the more complicated avoidance query, and learn to take advantage of their newly unlocked abilities. Here, we strengthen the position of our work by using the model’s processes to also account for a new study on second-order false-belief queries, where participants are asked about what another believes about a third party [1]. We show that our model can account for the new data with only procedural changes, strengthening its theoretical claims.

2. THEORY OF MIND MODEL
Our existing theory of mind process model hypothesizes that, for very simple ToM queries, such as the first-order FB query, the model relies on simple reasoning mechanisms (e.g., inhibiting their own, true, belief so that they can identify another’s belief [6]). The model also predicts that being able to select between beliefs in this way is a precursor for simulation, which allows people to use the beliefs and desires of others to predict and understand another’s behavior; this is ultimately what provides full-fledged ToM. Simulations have access to the model’s procedures and cognitive resources, and include the knowledge and beliefs needed to perform them. For the avoidance query, then, the model first predicts the other person’s belief like it does in the first-order FB task; then, it uses the identified belief as the basis for simulating the others’ actions to predict their behavior.

The model’s development is based on the idea that, as children grow, they learn and mature simultaneously; e.g., as they develop, they learn to take advantage of their maturing ability to select between competing beliefs. This occurs during two developmental shifts: the first where the model develops the ability to inhibit beliefs, and the second where the model develops the ability to perform simulation. Maturation is modeled via a developmental parameter; learning occurs via procedural utility learning.

The model fit the data very well for both a first-order FB task meta-analysis, and for an avoidance query experiment.
2.1 Model of Second-Order False Belief

Here, we extend our prior model to account for second-order false-belief queries. The process is identical to how the model answers avoidance queries: the model first uses inhibition to select a second person’s beliefs, and then uses that as the basis for simulation. The key difference is that the goal of simulation, instead of predicting another’s actions, is to identify a third person’s beliefs based on what the second person believes. For the second-order FB query, then, where the model is asked to identify someone else’s belief about a third person’s belief, the model identifies the second person’s belief, then spawns a simulation of the second person that, in turn, inhibits the second person’s belief to identify the third person’s belief.

The model develops this ability as it develops the ability to answer the avoidance query: the ability to perform simulation matures as the model simultaneously learns to use it to answer these types of queries.

4. REFERENCES


