The New School Psychology Bulletin
Volume 5, No. 1, 2007

Thinking About Me, You, and Them: Understanding Higher-Order Propositional Attitudes

Jason A. Wheeler Vega, Ph.D.[1]

Abstract ~ Higher-order propositional attitudes (HOPAs), such as "I think that you think that…," figure in many fields including theory of mind, cognitive ethology, and psycholinguistics. Analysis of many examples suggests there may be differences in understandability of HOPAs depending on the type and number of constituents and the presence of recursions. Empirical work on normal adult ability with HOPAs has been lacking, leaving research with special populations without a standard for comparison. An experiment explored the effects of varying the number of individuals in HOPA sentences, up to the eighth order. Significant differences in understandability of HOPA sentences were found between three groups, those that are about (i) oneself, (ii) dyads, and (iii) series of different individuals.

Early in *The Lion in Winter*, Goldman's (1966) play of politics, love, and language, the rebellious Eleanor of Aquitaine plots against King Henry II. Henry favors his youngest son, John, for

---

1: St. Luke's-Roosevelt Hospital Center, New York, NY, USA

Address correspondence to Jason Wheeler Vega: jasonwheelerphd@gmail.com

his succession. Eleanor favors their eldest, Richard. No one favors the middle child, Geoffrey. Eleanor schemes with Geoffrey, remarking that Henry still has his hopes set on John. Geoffrey replies: "I know. You know I know. I know you know I know, we know that Henry knows and Henry knows we know it. We're a knowledgable family." (p. 25; this exchange is also quoted by McGeer, 2001, to make a different point). Whatever dramatic effect this has on us, we are able to follow the content, though with some effort. Why can we understand utterances like this? And why, just when we feel stretched by the third round of "knowing" - "I know you know I know" - does it suddenly shift and seem to get easier - "we know that Henry knows and…."?

This kind of political psychology may come quite naturally to us; indeed, such dealings may have been a spur for human cognitive development (Byrne & Whiten, 1988; Dunbar, 1996; Humphrey, 1983). It should be no surprise that such sequences figure centrally in diverse fields. What is surprising is how little integrated and systematic theory and research there is on sequences like these, and that adult norms are not available to researchers for comparison with special populations. However, I will argue that one should not expect to find a single normative limit for sentences like these, but multiple limits, and that researchers have been misled by the formal similarity of such sentences into assuming a functional similarity. Analysis of published examples and predictions based on prior work will be discussed. I will then present an experiment with a college sample that examines adult understanding of sentences of three different kinds: (i) attitudes about one's own attitudes, (ii) attitudes that recur between dyads, and (iii) attitudes about several other people's attitudes. Finally, I will consider explanations for these differences and avenues for further research.

A Simple Analytical Scheme for HOPAs

People explain and predict what others will do by attributing to them mental states, including propositional attitudes such as belief, desire, hope, fear, and so on (Russell, 1921, 1940; Dennett, 1978a). Propositional attitudes can be about other propositional attitudes as well as directly about propositions, and here I'll call this a higher-order propositional attitude (HOPA). Even though one may use anything from propositional logic to more complex epistemic logics (Lismont & Mongin, 1994), here we do not need much more than the scheme Russell (1940, p. 210) gave us to analyze HOPAs: "A believes p."[1]

Consider the example from *The Lion in Winter*:

(1) I know. You know I know. I know you know I know, we know that Henry knows and Henry knows we know it.

A **k** *p*. B **k** A **k** *p*. A **k** B **k** A **k** *p*, A&B **k** C **k** A&B **k** *p*.

In (1) *I* (Geoffrey) is replaced by the individual variable A; *you* (Eleanor) by B; *know* is replaced by the propositional attitude variable **k** (other letters may represent other attitudes); and the *proposition* or content-sentence (elliptical in the original - "Henry still has his hopes set on John") is represented by the propositional variable *p*. One first notices that after three steps up the ladder in line, Geoffrey and Eleanor join forces (A&B), and a new individual, Henry, is introduced. Why this change? I will attempt to answer that question in the coming sections.

In this article I will be proposing that, although the variations are endless, it is most useful to consider three basic schemes for HOPAs that have a single proposition: *monadic* - A **k** A **k** *p* - with

[1]In the following analyses I will be assuming a matrix-sentence semantics for HOPAs, taken as embedded that-clauses (e.g., Harman, 1972; Larson & Ludlow, 1993/1997; and compare Davidson, 1984; Schiffer, 1987/1997; Burge, 1986; Lepore & Loewer, 1990).

one individual and recursions to that individual; *dyadic* **-** A **k** B **k** A **k** *p* **-** with recursions between only two individuals; and *polyadic* **-** A **k** B **k** C **k** D k *p* **-** in which higher orders develop but without recursions between individuals. These are taken as distinct forms as it is possible to theorize different functions for understanding each.

In line with a central assumption of evolutionary psychology (Barkow, Cosmides, & Tooby, 1992), it is worth considering that rather than being second, third, fourth, fifth, or n^{th} order intentional systems, *simpliciter*, it may be that humans have specific domains of functioning with HOPAs depending on whether they are thinking about themselves, one other agent, or a series of different agents. In other words, one might look for domain-specific cognitive adaptations rather than global cognitive functions. Researchers concerned with HOPAs have tended to assume the latter view, likely taking the common form of the propositional attitudes as defining a common cognitive function. On the other hand, one might instead look for what functions make use of the constituents of propositional attitudes and their iterations and combinations.

As regards monadic HOPAs, knowing one's own mind can be seen mainly as a benefit to the extent that it "permits our hypotheses to die in our stead" as Popper said (and Dennett elaborated, 1995, p. 375): inserting an occasion for choice between impulse and action. Beyond the second order of intention the practical value of further iterations is unclear. Concerning dyadic HOPAs, whereas it can be important to track what another knows about one's knowledge, long chains of reasoning may be easily interrupted, and, further, there is some argument for a profoundly random element in human thought and action (Miller, 1997). The value of repeated recursions between individuals is therefore questionable. With regard to polyadic HOPAs, in social contexts it is likely an advantage to be able to follow multiple relationships, in some cases making inferences along the way, to track who has been scratching whose back (and whose front), and what predictions follow from series of encounters. New individuals bring new "value," for want of a better term, to a HOPA, with new possible benefits and costs associated with each actor.

We can imagine HOPAs that contain mixtures of these elements: individuals and recursions, along with different attitude verbs, additional propositions, and so on. More complex examples are typical. For example, Dunbar theorizes that the great step in social intelligence between humans and even the cleverest of the other primates is our ability to understand fourth-order propositional attitudes (1996, p. 102, my brackets):

(2) "both the writer [A] and the reader [B] understanding what one character [C] thinks another character [D] wants the first character [C] to believe"

A&B **u** C **t** D w C **b** *p*

This is indeed a fourth-order HOPA, but contains four individuals, four different attitudes, and only a third-order dyadic recursion between C and D. This example is typical of writing and research using HOPAs, in that Dunbar focuses on only the order of embedding, and not on the constituents or recursions in the sentences. Other examples of the way theorizing is limited in this way will be discussed next.

Selective Survey of Literature on HOPAs

It will be seen in the following survey that theorists and experimenters have been hampered by the absence of a clear scheme and normative standards of adult ability to compare to their special populations. HOPAs appear most saliently for psychologists in the theory of mind literature. The now classic theory of mind experiments were inspired by cognitive ethology (Premack &

Woodruff, 1978; Dennett, 1978b). Wimmer and Perner (1983; Perner & Wimmer, 1985, 1987), produced a now classic series of experimental studies of false-belief and understanding of HOPAs, testing attribution up to the second order. Earlier work using non-behavioral tasks (Miller, Kessel, & Flavell, 1970; Eliot, Lovell, Dayton, & McGrady, 1979) tested ability to attribute second-order and third-order HOPAs (see also Schultz & Cloghesy, 1981). It is notable that the third-order HOPAs tested were polyadic and non-recursive: "The boy is thinking that the girl is thinking of the father thinking of the mother."

Early studies of normal adult understanding of HOPAs in the theory of mind literature made claims without evidence to have tested up to the fourth order, presenting in their reports neither naturally generated utterances nor findings about difficulty or success with different orders (Friedell, 1969; Laing, Phillipson, & Lee, 1966; Maucorps & Bassoul, 1962). A more recent "advanced test of theory of mind," used with adults (Happé, 1994) and older adults (Happé, Winner, & Brownell, 1998), employed story comprehension tasks to test second- and third-order intention understanding. "Another advanced theory of mind test," devised by Baron-Cohen and colleagues (Baron-Cohen, Jolliffe, Mortimore, & Robertson, 1997; Baron-Cohen, Wheelwright, Stone, & Rutherford, 1999) had adults infer mental states from photographs of people's eyes, demonstrating at most second-order understanding.

Clinicians and researchers have drawn on the ideas and research generated in the field of theory of mind, and training programs have been developed for autistic children aged 4 to 13 (Howlin, Baron-Cohen, & Hadwin, 1999; Swettenham, 2000; see also Leslie, 1991 and Baron-Cohen, 1995). But, unusually, these remedial training programs have no adult ability levels to aim for.

Deficits in theory of mind have been investigated in disorders other than autism, including schizophrenia, severe personality disorders, neurological disorders, conduct disorder, attention-deficit hyperactivity disorder, congenital deafness and blindness, with limited findings (Corcoran, 2000). These fields could all benefit from a conceptual scheme for understanding HOPAs. For example, Frith and Corcoran (1996) tested the ability of people with schizophrenia to infer the polite request, "Please shut the window," from the assertion, "It's very cold in here." But this more closely resembles a test of understanding illocutionary force (Austin, 1975), or inferring speech acts (Searle, 1969), rather than facility with HOPAs.

Psychoanalytic writers have also looked at HOPAs, under the names mentalization and reflective functioning (Fonagy & Target, 1998; Fonagy, Gergely, Jurist, & Target, 2002), and mirroring (Modell, 1984). They have suggested that being able to attribute mental states is a key psychological resource that helps individuals manage their emotions and behavior, and may predict progress in psychotherapy. While potentially valuable to clinicians, it is notable that this theory does not have normal ability levels with which to compare their special groups.

Computer scientists have also worked on modeling systems that use HOPAs (e.g., Wilks & Bien, 1983; Rapaport, 1986; Schultz, 1991), but have not paid attention to the elements of HOPAs, nor looked at recursions or very long iterations. For instance, Taylor, Carletta, and Mellish (1996) argue that deceptive communication requires systems to process fourth-order, or even higher, intentions. They give a borrowed example: SBUBSBFDS [System believes User believes System believes Frank dislikes System]. But this example is ambiguous and could be a fourth-order HOPA with three individuals and two attitudes, or a third-order dyadic recursive HOPA whose proposition is *Frank dislikes System*. Both interpretations qualify their claim. HOPAs have also been a subject of research on Prisoner's-Dilemma-type games, and players have been found to attribute second-, but not third-order, dyadic HOPAs to competitors; higher orders were not tested (Hedden & Zhang, 2002).

Philosophers have speculated about whether our attributions of propositional attitudes to others may be extended indefinitely or not (Radford, 1969; Cargile, 1970; Dummett, 1986). Rosenthal (1986, 1993a, 1993b, 2002) has developed a theory of consciousness which implies that being conscious of an introspected mental state would be a fourth-order monadic state. While some of these writers have considered practical constraints on attributing HOPAs, none have had empirical work to draw upon or attended to the possible effects of different constituents or recursions on understanding of HOPAs.

Among contemporary philosophers, Dennett has proposed two of the most interesting hypotheses about understanding HOPAs. First, he suggests that it is the step from first-order to second-order that is crucial. Facility within the domain of HOPAs is just a matter of memory, "so that there seems no interesting difference between, say, a fourth-order and a fifth-order intentional system." (1987, pp. 244-5). He guesses humans can understand about five or six orders, at best (ibid., p. 243). However, it is plausible that polyadic HOPAs of higher than sixth order might be quite understandable to normal adults, and that lower order recursive monadic and dyadic HOPAs might, on the other hand, not be. Talking about an intentional system's ability, singular, may therefore be restrictive. Second, Dennett (1996, brackets added) has also suggested that context is responsible for comprehension of HOPAs:

> Sometimes higher orders are so easy as to be involuntary. Why is this fellow in the movie trying so hard to avoid smiling? In the context it's deliciously obvious: his effort shows us [3] he knows she doesn't realize he already knows she

> wants him to ask her to the dance, and he wants to keep it that way! Other times, simpler iterations can stump us. [4] Are you sure that I want you to believe that I want you to believe what I'm saying here? (p. 121)

Examine his examples:

(3) A **b** B ~**b** A **b** B **d** *p*, AND A **d** *q* OR A **d** B ~**b** A **b** B **d** *p*

This is an ambiguous, at most fourth-order dyadic HOPA that also uses negation and different propositional attitudes.

(4) A **b** B **d** A **b** B **d** A **b** *p* OR A ~**b** B **d** A **b** B **d** A **b** *p*

His second example is a fifth-order dyadic HOPA, ambiguous at the main verb, which alternates belief and desire verbs. If (4) is harder to understand than (3), is that because of a lack of context, or because of differences in the constituents, or because (4) contains more than two recursions per individual? Two major hypotheses about the understandability of HOPAs proposed by Dennett are not supported by analysis of the form and constituents of his examples. It may be, however, that there is something like "context" which influences understandability of HOPAs.

A rich model of the role of context is found in psycholinguistic work by Clark and Marshall (1981/1992) on mutual-knowledge (common-knowledge, Lewis, 1969).[2] Looking at the development of shared-knowledge in dyads, Clark and Marshall tell sto-

[2]Psycholinguistic work on two kinds of metacognition, referred to as feeling-of-knowing and feeling-of-another's-knowing, has examined second-order attributions of propositional attitudes (Brennan & Williams, 1995). This research, as the names suggest, concerns subjective confidence about one's own (monadic) or another's (dyadic) beliefs, but does not use objective measures of ability with HOPAs.

ries about two characters and what they know about which movie is playing at a local theater. They construct increasingly complex scenarios. Their first example involves only two characters (individuals) and one proposition (the film title). Longer scenarios are constructed, leading to their fifth order scenario, summarized as follows: A and B read the paper and together learn that *R**. Then B reads a late edition and learns that not *R** but R, and marks the correction in this second edition. Then A reads the second edition and notices B's marking of the correction. B sees A notice this. In a mirror, A sees B seeing her notice this, but realizes that B did not notice her see him in the mirror seeing her notice his mark.

Clark and Marshall (1981/1992) claim: "In principle, we could use this procedure to construct countermanding versions ad infinitum" (p. 14). However, their examples actually illustrate how difficult it is to devise scenarios for recursive dyadic HOPAs. For instance, they have to introduce two propositions and separation of the actors in space and time, along with objects (papers) to record changes of proposition. And, perhaps most tellingly, they do not go on, though they say that they might. The assumption that further versions could be constructed ad infinitum rests on the fact that it is possible to recursively extend a HOPA *sentence*. This is not the same thing, however, as constructing an *enactable scenario* for a HOPA sentence.

If dyads don't process shared-knowledge attributions to very many orders, they may instead, Clark and Marshall (1981/1992) suggest, use various *copresence heuristics* for establishing direct reference. These vary in the quality of the grounds they provide for the speaker's use of definite descriptions and may require assumptions based on community membership, physical or linguistic copresence, or mixtures of these. The strongest grounds are given by the *immediate physical copresence* of speaker, addressee, and referent. The less direct the links between these three variables the more assumptions are needed. Turning this observation around, while immediate physical copresence may require the least cogitation, it is possible that less direct situations may both require and allow HOPA sentences to become longer while remaining understandable. So, for example, Radford (1969; see also Cargile, 1970) tells a story about two people that seems to build up an indefinite (recursive) dyadic HOPA. But he does so by separating the actors in space and time, and introducing things like writing in a diary, which is then read later, and people overhearing things through walls. Though the resulting HOPA sentence would be the same, there may be important differences in understandability of HOPAs depending on the co-presence relations between the individuals involved.

The foregoing review has shown that many fields make essential use of understanding HOPAs. There are different intuitions about understanding HOPAs, and no comprehensive research that explores adult ability which would establish standards for comparison for researchers. It is the main conclusion of this review, and the chief argument of this article, that in exploring adult ability with HOPAs one is more likely to discover that there are multiple limits for adult understanding of HOPAs - in line with evolutionary assumptions - and not a single level, depending on features of the HOPAs concerned. These may include the number of individuals, recursions between individuals, and co-presence relations between individuals in the development of an occasion for the use of a HOPA, amongst other variables.

In terms of properties of HOPAs that affect understanding, particularly relevant is the role of recursion. The assumption that recursions may be applied indefinitely to the same individuals is challenged by many counterexamples that show unrecognized limits at which people add elements to HOPAs, such as new individuals or other constituents - as in Goldman's example (1) - in order to limit recursions. Prior attempts at formalizing the relationship between recursion and understanding have been made, but without clear results. Clark and Marshall (1981/1992) distinguish two kinds of recursion: *pipeline* recursion (dyadic) and *reciprocal* recursion (polyadic). They say that the difficulty of understanding HOPAs increases with the square of the number of reciprocal recursions. But this difficulty factor is left unanalyzed, nor are limits considered. Similarly, Yngve's (1960) concept of depth does not make differential predictions between monadic, dyadic, and polyadic forms, as in his scheme all progressive, or right branching, constructions have an *Yngve depth* of 1. He also suggested an ungrammatical first step rule for limiting depth. But this would make second-order monadic HOPAs, like "I know that I know that *p*," ungrammatical at the first recurrence of "I know that," which is counterintuitive.

With these considerations in mind, and from the analysis of many HOPAs in various literatures, it is hypothesized that understanding of HOPA sentences will depend not just on the length of the HOPAs, but on their constituents, number of recursions, and type of co-presence of the actors. Assuming a situation of physical co-presence, it is hypothesized that HOPA sentences with a single proposition and containing only one type of propositional attitude verb will be understandable as long as they contain no more than two recursions per individual. Therefore: monadic recursive sentences should be understandable at up to the second order; dyadic recursive sentences should be understandable at up to the fourth order; and polyadic nonrecursive sentences should be understandable to the eighth order (highest tested), and ultimately to the limits of long-term memory.

Method

Developing an appropriate, exploratory methodology for an empirical study of HOPAs was a significant challenge. One variable was chosen for this preliminary empirical study: the number of individuals (from one to eight) in sentences of up to the eighth order; this is higher than tested in previous research. A single propositional attitude, knowing, was chosen to simplify interpretation of the results. Several methods were considered and rejected because they limited the range of responses, or constrained the variety of co-presence (Clark & Marshall, 1981/1992) in the experiment, where a broadly exploratory method was desired for this study. Rejected options included methods related to Wimmer and Perner's (1983) false-belief task; true-false responding to experimenter-constructed scenarios; and a simulated email program with iterated delivery reply notices. With this exploratory goal in mind, a method was chosen that would allow participants both to *say* whether they understood the sentences, and would also require them to *generate* scenarios for the stimulus sentences that would demonstrate their understanding. This combination test of understanding includes both comprehension and production tasks. Further, a mixture of questionnaire and protocol analysis/think aloud methods were used to allow the addition of qualitative data (Ericsson & Simon, 1980; Ericsson & Crutcher, 1991). Both concurrent and retrospective think-aloud data (Fonteyn, Kuipers, & Grobe, 1993; Kuusela & Paul, 2000) were recorded on audiotape, and later transcribed verbatim.

Participants

Thirty volunteers took part (21 women and 9 men, mean age = 23.3 years, range 19 to 39 years). Twenty-nine were students (21 undergraduate, 8 graduate) and one was an employee of New School University. Half participated for research credit through the New School Psychology Subject Pool and half for monetary compensation. Participants described themselves as: Caucasian or White (22), Asian (2), Eurasian (2), Hispanic (2), Black (1), and South Asian (1). They spoke between one and seven languages. Most (22) were monolingual English speakers, though ten of these reported some second language study in high school or college. Five participants were bilingual in English and another language (Hebrew, Italian, Korean, Spanish, Tamil) from birth.

Three participants' first languages were not English (French, Spanish, Tibetan), but these participants had learned English in primary (2) or secondary school and were considered fluent in English by themselves and by the experimenter. (Language experience did not affect the results of the experiment.)

Materials

Participants responded to materials developed by the author. The materials were three sets of HOPA sentences - monadic, dyadic, and polyadic. Each set comprised eight sentences, beginning with a first-order sentence and continuing to the eighth order. Twenty participants completed the questionnaire-only method and 10 completed the think-aloud method. The question sets were the same for both methods (Appendix A), and were piloted with graduate students from the New School Psychology Department.

Design and Procedure

Each participant responded to each of three sets of HOPA sentences; order of sets was counterbalanced. Participants were randomly assigned to questionnaire or think-aloud method. Participants were given the materials in a stapled packet. The instructions were read aloud to the participant while the participant read them also. Questions about the instructions were invited. For questionnaire participants the experimenter then left the room, instructing the participant to signal when finished. For think-aloud participants the experimenter remained in the room and began audio recording after the instructions were given (Appendix A; instructions for both conditions were nearly identical), and asked open probe questions about participant's remarks on the sentences: "Tell me more about…?" This procedure took between 10 and 20 minutes per participant.

Scoring

Questionnaire and think-aloud methods provided several items of data, which required different scoring and analysis procedures:

first, the highest number claimed (circled) for each question set was recorded; next, it was noted for which sentence the participant actually gave an example; third, the situation examples given were scored for order (third-order, seventh-order, etc.), and finally reports of why sentences were or were not understandable were noted.

The scoring was conducted using the analytic scheme outlined earlier, along with principles developed during pilot scoring (see Appendix B). These principles aimed to provide both definite values suitable for quantitative analysis, and conservative scoring of examples that might be taken as implying infinite regressions - these would confound quantitative comparisons - as in scenarios that refer to common knowledge among immediately co-present agents (Clark & Marshall, 1981/1992).

Results

Overall effects were examined with analysis of variance (ANOVA) for repeated measures and group differences examined with contrasts. There was a significant difference between *claims* made about the three forms of HOPA, $F(2, 58) = 17.18$, $p < .001$. Participants claimed to understand polyads at higher orders than dyads, and dyads at higher levels than monads, linear contrast $F(1, 29) = 26.77$, $p < .001$. There was no main or interaction effect (alpha .05) in this study of any of the participant variables listed above.

There was also a significant difference between the three sentence schemes by the order at which examples were *attempted*, $F(2, 58) = 36.11$, $p < .001$. Participants attempted to give examples for polyads at higher levels than dyads, and dyads at higher levels than monads, linear contrast $F(1, 29) = 61.91$, $p < .001$. There were no other main or interaction effects. Orders claimed and attempted do not differ logically, but do show that making an attempt at exemplifying higher orders was less frequent (likely harder) than simply claiming to understand them.

Again, there was a significant difference between the three sentence schemes depending on the highest-order sentences *scored*, $F(2, 58) = 50.62$, $p < .001$. Participants were able to give examples for polyads at a significantly higher order than for either monads or dyads, quadratic contrast $F(1, 29) = 15.45$, $p < .001$. The scored level of monadic examples did not differ significantly between think-aloud and questionnaire participants, but think-aloud participants gave significantly higher level examples than questionnaire participants for both dyadic and polyadic sentences, $F(1, 28) = 12.31$, $p = .002$. There were no other main or interaction effects. Table 1 shows means and standard deviations used in these analyses.

Close analysis of HOPAs in many literatures suggested that HOPAs are not homogeneous, as usually assumed, and that manipulations of their consitutents would produce schemes that were differently understandable. The aims of this experiment were to test this prediction and to explore the limits of normal adult ability with three different forms of HOPA sentences. Participants found polyadic HOPA sentences - containing many different individuals, with no recursions - easier to understand at higher orders than those containing one or two individuals with recursions. In turn, they found recursive sentences containing two individuals easier to understand at higher orders than recursive sentences containing only one individual. Although there were significant differences in the predicted order between the three kinds of HOPA sentence, the mean levels of knowing did not precisely match the specific predictions (Table 1). For both monads and dyads, most participants claimed to understand higher levels than hypothesized, but most were only able to give examples for lower levels than hypothesized. For polyads, however, participants' mean claimed understanding was only just below the predicted level.

Table 1

Mean Higher-Order Propositional Attitude Understanding

	Sentence type			
	Claimed	Attempted	Scored	Predicted Maximum
Monadic				
M	4.60	3.53	1.10	2
SD	2.81	2.45	.96	
Dyadic				
M	5.33	4.67	1.93	4
SD	1.77	1.52	1.70	
Polyadic				
M	7.40	7.30	5.97	8
SD	1.38	1.64	2.83	

Note. *n* = 30.

Qualitative data were also examined. Here is a typical first-order monadic scenario: "I put the chocolate in the green cupboard, therefore I know that it's there." The highest scored examples given for monadic sentences were third order. Interestingly, two were of academic discussions of metacognition. For example: “When discussing consciousness, maybe in a philosophy class” and “When you want to assert that you are aware of your knowledge of knowing.”

The highest dyadic interpretation given was eighth order. The example resembled iterated knowledge stories by Radford (1969) and Cargile (1970), and involved *linguistic*-, not *physical*-, *copresence*, with knowledge passed between the actors by two other parties. The participant's example is:

> Somebody told him and then they told me, so I would know if he knows, and then Patrick knows that I know that because maybe he spoke to that person, and maybe I would know that Patrick knows that because I'd spoken to somebody else who overheard them. And maybe he would know that I'd know because that person told *him*, and then I would know that because maybe that same person told me the whole situation . . . which I guess would mean that number 8 would also make sense, but it's just even further and deeper down the line.

Many polyadic examples described individuals passing information one to the other by word of mouth, as for instance: "Well, having seen someone put the chocolate in the cupboard, if I went and then told Alex that I knew that it was in the cupboard and she said, 'Well, I'm not gonna go take it cause I'm not hungry,' but she told Peter, so Peter knows that Alexa knows. Peter then goes and tells Anna…"

A number of scenarios given for the eighth-order polyadic sentences failed to meet coding principle 3 (Appendix B) by grouping individuals equivalently instead of reflecting the differences in their states of knowledge, indicated by their place in the sentences, such as examples in which all actors were present and observed the chocolate being placed in the cupboard.

Discussion

In exploring normal adult understanding of HOPAs in a college sample one might have expected, as previous researchers have assumed, that there is a single level of ability that increases with age and reaches a ceiling. It is presently known that around 3 years of age humans can be attributed some part-concept of belief (Bennett, 1991; Wellman, 1991), which one might call egocentric belief. It is further known that the ability to attribute false-beliefs develops around age 4, and this is widely taken as a "watershed" (Astington, 1991, p. 159) at which children acquire a full belief-desire psychology and theory of mind. It is also known that from around age 7 children can attribute, as well as be attributed, second-order dyadic attitudes (Perner & Wimmer, 1985). Using the distinctions made in this article, one may distinguish this achievement from the ability to attribute second-order monadic attitudes, which may be rare before 9 years of age (Miller et al., 1970). Based on this experiment, it is not yet possible to add definitive normative levels of adult ability to this developmental sequence. However, this study has shown that looking for one level of ability is likely to be a mistake, and that such factors as number of individuals, recursions, and copresence may determine the understandability of HOPAs. In addition, in this experiment the opportunity for participants to think-aloud and revise their examples through talking about them may have improved their performance on the task. This methodological effect suggests that further research on HOPAs should attend to this factor in understanding HOPAs.

Monadic HOPAs were understandable only at relatively low levels, at most third order. An educative function beyond the second order, communicating the value of metacognition (second order) to others, is perhaps the practical limit for monadic HOPA use. With regard to theorized functions of dyadic HOPAs, it is notable that a third of participants (10) in the experiment gave examples that involved some element of deception or conflict. Why would a psychological "arms race" between deception and counter-deception not have made us all such that in dyadic interactions people regularly attribute eighth-order, or higher, intentions? It is possible that in dyadic interactions, as suggested by analyzed examples and the experiment presented here, people may climb the ladder of HOPAs at most to the third order, and then switch to an assumption of *common knowledge* (Lewis, 1969). Or, the ladder may be ducked altogether: The economical co-presence heuristics for referring suggested by Clark and Marshall (1981/1992) shortcut potentially endless conditions for agreeing when something is mutually known, as do the zero-order algorithms most successful in Prisoner's Dilemma games. Also, as Grice (1969/1997) argues, in the absence of behavioral cues to differentiate levels of HOPAs, at some early, unspecified stage a listener will cease to have intentions about a speaker's intentions, and vice versa. Humans perhaps followed the evolutionary path of refining the ability to attribute attitudes accurately at lower levels, instead of the path of adding levels of attribution. Whereas different individuals, as in polyadic HOPAs, are worth tracking for their unique social effects, a capricious move on the part of a dyadic actor may upset a long and effortful chain of dyadic inference. Non-recursive, polyadic HOPAs might be useful to the limits of long-term memory, though further study should be made of possible applications for long polyadic HOPAs.

This analysis and these findings have a number of implications for fields that are directly or indirectly concerned with HOPAs. First, these fields would benefit from no longer treating HOPAs as a homogeneous group underwritten by a global cognitive function, but instead paying attention to specific cognitive abilities for particular intra- and inter-personal functions and accompanying variations in their constituents. Many variables remain to be explored empirically. For example, altering the *tense* of an attitude verb might increase understanding of recursive HOPAs, as in: I thought that I shall think that I think that I think that the chocolate is in the green cupboard. Similarly, including *different attitudes* might increase understanding. For example, one participant said of a third monadic order sentence, "When you want to *assert* that you are *aware* of your *knowledge* of knowing". There are many possible variations, such as: I think that you want that I hope that you fear that I expect that you wish that the chocolate is in the green cupboard. Other variables were seen at work in some examples analyzed earlier, such as negation, ambiguity, and additional propositions, which might also increase understanding by enriching the basic schemes.

Within different fields some factors will be more relevant than others. So, for example, clinicians might focus on helping people to more consistently make the 'crucial second step' (Dennet, 1987, p. 243) towards accurately attributing attitudes to themselves and others. A line of research of particular interest concerns the proposed limitations of people diagnosed with severe personality disorders for accomplishing what has been called mentalizing (Fonagy & Target, 1998), which in the terms of this article may be the ability to attribute second- or third-order dyadic HOPAs. Interesting questions in this domain include whether this is a consistent or temporary - trait or state - deficit, and whether people might improve their abilities with HOPA attribution as a result of various psychotherapies. Such questions could be addressed using the methodology applied here, or with other methods that are based in the same analysis.

In conclusion, I have presented a simple analytic framework for HOPAs, an evolutionary-modular conceptual framework, preliminary empirical evidence, and a quantified role for recursion as a factor limiting understanding, which together suggest distinct patterns of adult understanding of HOPAs that may reflect discrete adaptive purposes for these abilities. But, as is familiar, the telling of a plausible evolutionary story should be supported by special evidence of adaptation (Williams, 1966). These experiments have not provided such evidence. It is possible, for example, that English imposes grammatical structures that affect HOPAs which are absent or different in other languages. However, the cultural and linguistic diversity of the participants in this study and lack of effects for these participant variables is at least not discouraging in this regard. Further research on understanding HOPAs, including work in other languages, may extend understanding of the functions for which HOPAs are used.

References

Astington, J. W. (1991). Intention in the child's theory of mind. In D. Frye & C. Moore (Eds.), *Children's theories of mind: Mental states and social understanding* (pp. 157-172). Hillsdale, NJ: Lawrence Erlbaum Associates.

Austin, J. L. (1975). *How to do things with words* (2nd ed.). Cambridge, MA: Harvard University Press.

Barkow, J. H., Cosmides, L., & Tooby, J., (1992). *The adapted mind: Evolutionary psychology and the generation of culture*. Oxford: Oxford University Press.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA: Bradford.

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with autism or Asperger syndrome. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 38*(7), 813-822.

Baron-Cohen, S., Wheelwright, S., Stone, V., & Rutherford, M. (1999). A mathematician, a physicist and a computer scientist with Asperger syndrome: Performance on folk psychology and folk physics tests. *Neurocase: Case Studies in Neuropsychology, Neuropsychiatry, and Behavioural Neurology, 5*(6), 475-483.

Bennett, J. (1991). How to read mind in behavior: a suggestion from a philosopher. In A. Whiten (Ed.), *Natural theories of mind: Evolution, development and simulation of everyday mindreading* (pp. 97-108). Oxford: Blackwell.

Brennan, S. E., & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language, 34*, 383-398.

Burge, T. (1986). On Davidson's "saying that". In E. LePore (Ed.), *Truth and interpretation: Perspectives on the philosophy of Donald Davidson* (pp. 190-208). Oxford: Blackwell.

Byrne, R. W., & Whiten, A. (1988). *Machiavellian intelligence: Social expertise and the evolution of intellect in monkeys, apes, and humans*. Oxford: Clarendon.

Cargile, J. (1970). A note on "iterated knowings." *Analysis, 30*, 151-155.

Clark, H. H., & Marshall, C. R. (1992). Definite reference and mutual knowledge. In H. H. Clark (Ed.), *Arenas of language use* (pp. 9-59). Chicago: University of Chicago Press. (Original work published 1981)

Corcoran, R. (2000). Theory of mind in other clinical conditions: is a selective 'theory of mind' deficit exclusive to autism? In S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds.), *Understanding other minds: Perspectives from cognitive neuroscience* (2nd ed., pp. 391-421). Oxford: Oxford University Press.

Davidson, D. (1984). *Inquiries into truth and interpretation.* Oxford: Clarendon.

Dennett, D. C. (1978a). *Brainstorms: Philosophical essays on mind and psychology.* Cambridge, MA.: MIT Press.

Dennett, D. C. (1978b). Beliefs about beliefs (commentary). *Behavioral and Brain Sciences, 1*, 568-570.

Dennett, D. C. (1987). *The intentional stance*. Cambridge, MA: MIT Press.

Dennett, D. C. (1995), *Darwin's dangerous idea: Evolution and the meanings of life*. New York: Simon and Schuster.

Dennett, D. C. (1996). *Kinds of minds: Toward an understanding of consciousness.* New York: Basic Books.

Dummett, M. (1986). A nice derangement of epitaphs: Some comments on Davidson and Hacking. In E. LePore (Ed.), *Truth and interpretation: Perspectives on the philosophy of*

Donald Davidson (pp. 459-76). Oxford: Blackwell.

Dunbar, R. (1996). *Grooming, gossip, and the evolution of language.* Cambridge, MA: Harvard University Press.

Eliot, J., Lovell, K., Dayton, C. M., & McGrady, B. F. (1979). A further investigation of children's understanding of recursive thinking. *Journal of Experimental Child Psychology, 28,* 149-157.

Ericsson, K. A., & Crutcher, R. J. (1991). Introspection and verbal reports on cognitive processes - two approaches to the study of thinking: A response to Howe. *New Ideas in Psychology, 9,* 57-71.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87,* 215-251.

Fonagy, P., & Target, M. (1998). Mentalization and the changing aims of child psychoanalysis. *Psychoanalytic Dialogues, 8,* 87-114.

Fonagy, P., Gergely, G., Jurist, E. L., & Target, M. (2002). *Affect regulation, mentalization, and the development of the self.* New York: Other Press.

Fonteyn, M. E., Kuipers, B., & Grobe, S. J. (1993). A description of think aloud method and protocol analysis. *Qualitative Health Research, 3,* 430-441.

Friedell, M. F. (1969). On the structure of shared awareness. *Behavioral Science, 14*(1), 28-39.

Frith, C. D., & Corcoran, R. (1996). Exploring 'theory of mind' in people with schizophrenia. *Psychological Medicine, 26*(3), 521-30.

Goldman, J. (1966). *The lion in winter.* New York: Samuel French.

Grice, H. P. (1997). Utterer's meaning and intentions. In P. Ludlow (Ed.), *Readings in the philosophy of language* (pp. 59-88). Cambridge, MA: Bradford. (Original work published 1969)

Happé, F. G. E. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped and normal children and adults. *Journal of Autism and Developmental Disorders, 24,* 129-154.

Happé, F. G. E., Winner, E. & Brownwell, H. (1998). The getting of wisdom: Theory of mind in old age. *Developmental Psychology, 34*(2), 358-362.

Harman, G. (1972). Logical Form. *Foundations of Language, 9*(1), 38-65.

Hedden, T., & Zhang, J. (2002). What do you think I think you think? Strategic reasoning in matrix games. *Cognition, 85,* 1-36.

Howlin, P., Baron-Cohen, S., & Hadwin, J. (1999). *Teaching children with autism to mind-read: A practical guide for teachers and parents.* New York: John Wiley & Sons.

Humphrey, N. (1983). *Consciousness regained: Chapters in the development of mind.* Oxford: Oxford University Press.

Kuusela, H., & Paul, P. (2000). A comparison of concurrent and retrospective verbal protocol analysis. *American Journal of Psychology, 113,* 387-404.

Laing, R. D., Phillipson, H., & Lee, A. R. (1966). *Interpersonal perception: A theory and method of research.* London: Tavistock.

Larson, R. K., & Ludlow, P. (1997). Interpreted logical forms. In P. Ludlow (Ed.), *Readings in the philosophy of language* (pp. 993-1039). Cambridge, MA.: Bradford. (Original work published 1993)

Lepore, E., & Loewer, B. (1990). A study in comparative semantics. In C.A. Anderson & J. Owens (Eds.), *Propositional attitudes: The role of content in logic, language, and mind* (pp. 91-111). Stanford: Center for the Study of Language and Information.

Leslie, A. M. (1991). The theory of mind impairment in autism: evidence for a modular mechanism of development? In A. Whiten (Ed.), *Natural theories of mind: Evolution, development and simulation of everyday mindreading* (pp. 63-78).

Oxford: Blackwell.

Lewis, D. K. (1969). *Convention: A philosophical study*. Cambridge, MA: Harvard University Press.

Lismont, L., & Mongin, P. (1994). On the logic of common belief and common knowledge. *Theory and Decision, 37,* 75-106.

Maucorps, P. H., & Bassoul, R. (1962). Jeux de miroirs et sociologie de la connaissance d'autrui. [Games of mirrors and sociology of knowledge of others.]. *Cahiers Internationaux de Sociologie,* 32, 43-60.

McGeer, V. (2001). Psycho-practice, psycho-theory and the contrastive case of autism: How practices of mind become second-nature. *Journal of Consciousness Studies, 8,* 109-132.

Miller G. F. (1997). Protean primates: The evolution of adaptive unpredictability in competition and courtship. In A. Whiten & R. W. Byrne (Eds.), *Machiavellian intelligence II: Extensions and evaluations* (pp. 312-340). Cambridge: Cambridge University Press.

Miller, P. H., Kessel, F. S., & Flavell, J. H. (1970). Thinking about people thinking about people thinking about...: A study of social cognitive development. *Child Development, 41,* 613-23.

Modell, A. (1984). *Psychoanalysis in a new context*. New York: International University Press.

Perner, J., & Wimmer, H. (1985). "John thinks that Mary thinks that..." Attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology, 39,* 437-471.

Perner, J., & Wimmer, H. (1987). Young children's understanding of belief and communicative intention. *Pakistan Journal of Psychological Research, 2*(1-2), 17-40.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 4,* 515-26.

Radford, C. (1969). Knowing and telling. *The Philosophical Review, 78,* 326-336.

Rapaport, W. J. (1986). Logical foundations for belief representation. *Cognitive Science, 10,* 371-422.

Rosenthal, D. M. (1986). Two concepts of consciousness. *Philosophical Studies, 49,* 329-359.

Rosenthal, D. M. (1993a). Higher-order thoughts and the appendage theory of consciousness. *Philosophical Psychology, 6*(2), 155-166.

Rosenthal, D. M. (1993b). Thinking that one thinks. In M. Davies & G. W. Humphries (Eds.), *Consciousness: Psychological and Philosophical Essays* (pp. 197-223). Oxford: Blackwell.

Rosenthal, D. M. (2002). How many kinds of consciousness? *Consciousness and Cognition, 11,* 653-665.

Russell, B. (1921). *The analysis of mind.* New York: Macmillan.

Russell, B. (1940). *An inquiry into meaning and truth*. New York: W. W. Norton.

Searle J. R. (1969). *Speech acts*. Cambridge: Cambridge University Press.

Schiffer, S. (1997). Sententialist theories of belief. In P. Ludlow (Ed.), *Readings in the philosophy of language* (pp. 856-873). Cambridge, MA: Bradford. (Original work published 1987)

Schultz, T. R., & Cloghesy, K. (1981). Development of recursive awareness of intention. *Developmental Psychology, 17*(4), 465-471.

Schultz, T. R. (1991). Modelling embedded intention. In D. Frye & C. Moore (Eds.), *Children's theories of mind: Mental states and social understanding* (pp. 195-218). Hillsdale, NJ: Lawrence Erlbaum Associates.

Swettenham, J. (2000). Teaching theory of mind to individuals with autism. In S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds.), *Understanding other minds: Perspectives from cognitive neuroscience* (2nd ed., pp. 442-456). Oxford: Oxford University Press.

Taylor, J. A., Carletta, J., & Mellish, C. (1996). Requirements for belief models in cooperative dialogue. *User Modeling and User-Adapted Interaction, 6,* 23-68.

Wellman, H. M. (1991). From desires to beliefs: Acquisition of a theory of mind. In A. Whiten (Ed.), *Natural theories of mind: Evolution, development and simulation of everyday mindreading* (pp. 19-38). Oxford: Blackwell.

Wilks, Y., & Bien, J. (1983). Beliefs, points of view, and multiple environments. *Cognitive Science, 7,* 95-119.

Williams, G. C. (1966). *Adaptation and natural selection: A critique of some current evolutionary thought.* Princeton, NJ: Princeton University Press.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13,* 103-128.

Yngve, V.H. (1960). A model and an hypothesis for language structure. Proceedings of the American *Philosophical Society, 104*(5), 444-466.

Author Note

This article is based on a thesis completed as part of an MA in Psychology at the New School for Social Research. Many thanks to Professor Michael Schober for his support of and assistance with this work from the earliest stages. Thanks also to Professor Nick Humphrey, who offered directions to literary examples and to pertinent research, and suggested key early improvements to the materials. I am grateful to members of the clinical psychology doctoral program and the psycholinguistics lab group at The New School for Social Research, who helped pilot the experiment and provided useful discussion.

Appendix A

Think-aloud Instructions and Sentences

Please read each of the sentences on the following pages. They are divided into 3 sections, headed A, B, and C.

- Throughout this study, starting when you turn the page to begin the task, please constantly think aloud. If you stop thinking aloud for more than a few seconds, the experimenter will prompt you to keep thinking aloud. This means, as you are reading the sentences and trying to understand them, just say whatever you are thinking out loud. You may want to read the sentences out loud while you are doing this. You may want to talk aloud about any difficulty you have with the sentences or parts of the sentences. This will help us to understand how people go about understanding the sentences, and clarify any possible difficulties.

- There are 8 sentences in each section. Circle the number of each sentence, 1-8, that you understand.

- For any sentence that you don't understand, please say, in a few words, why not.

- In each of the 3 sections, for only the highest numbered sentence that you do understand, please think aloud an example of a real life situation in which that sentence might be used. See the **Situation Examples** given at the foot of the page in each section. If you can't give a **Situation** for the highest numbered sentence you circled, go back through the lower numbered sentences until you find one you can give a **Situation** for.

- After you've completed the task, the experimenter may ask you questions about some of the things you said while thinking aloud.

- You may refer back to these instructions at any time during the study.

Monadic Sentences

1 I know that the chocolate is in the green cupboard.

2 I know that I know that the chocolate is in the green cupboard.

…

8 I know that I know that I know that I know that I know that I know that I know that I know that the chocolate is in the green cupboard.

Dyadic Sentences

1 I know that the chocolate is in the green cupboard.

2 Patrick knows that I know that the chocolate is in the green cupboard.

…

8 Patrick knows that I know that Patrick knows that I know that Patrick knows that I know that Patrick knows that I know that the chocolate is in the green cupboard.

Polyadic Sentences

1 I know that the chocolate is in the green cupboard.

2 Alexa knows that I know that the chocolate is in the green cupboard.

…

8 Patrick knows that Laura knows that Eve knows that Michael knows that Anna knows that Peter knows that Alexa knows that I know that the chocolate is in the green cupboard.

Situation Examples: Here are some examples of a real life situation in which *Sentence 1* might be used: I watch someone else put the chocolate into the cupboard; I'm told by someone else that the chocolate is in the cupboard; I remember that yesterday I put the chocolate into the cupboard. For only the highest numbered sentence that you do understand, please give an example of a real life situation in which that sentence might be used.

Appendix B

Scoring Principles

(P1) Examples must describe specific, enactable situations.

(P2) Abnormal examples are coded zero-order.

(P3) Examples must reflect the logical form of sentences.

(P4) Common knowledge in polyads is coded as first-order.

(P5) Common knowledge in dyads is coded as third-order.