

The London School of Economics and Political Science

Developing the cyranoid method of mediated interpersonal communication in a social psychological context: Applications in person perception, human-computer interaction, and first-person research

Kevin C. Corti

A thesis submitted to the London School of Economics and Political Science's Department of Social Psychology for the degree of Doctor of Philosophy

November, 2015

DECLARATION

I certify that this thesis, presented for examination for MPhil/PhD degree at the London School of Economics and Political Science, is solely my own work except for where clearly indicated (in which case the extent of any work carried out jointly be me and any other person is clearly identified in it).

The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made. Any citation of the work in this thesis that has been published by an academic journal or other publisher must give credit in accordance with standard referencing practices. This thesis may not be reproduced without my prior written consent.

I warrant that this authorization does not, to the best of my belief, infringe the rights of any third party.

A handwritten signature in black ink, appearing to read "Kevin C. Corti".

Kevin C. Corti

ABSTRACT

This thesis revisits Stanley Milgram's "cyranoid method" of interactive social psychological experimentation (Milgram, 2010a) and explores the technique's empirical potential in several domains. The central component of the method is speech shadowing, a procedure that involves a person (the shadower) repeating in real-time words they receive through an inner-ear monitor by-way-of radio-relay from a remote source. Speech shadowing effectively creates a hybrid agent (a "cyranoid") composed of the body of one individual (the shadower) and the "mind" (or more precisely, the words) of another (the source). Interactants naïve to this manipulation perceive speech shadowers as autonomous communicators, and this perceptual bias (the "cyranic illusion") affords researchers the ability to inspect the effects of separately altering the physical (outer) and dispositional (inner) elements of an interlocutor's identity in contexts involving spontaneous and unscripted face-to-face dialog.

Four articles and two additional chapters have been developed for this thesis. Chapter 1, "Introducing and situating the cyranoid method" presents an overview of the cyranoid method alongside an analysis of documents pertaining to the method contained in the Stanley Milgram Papers archive at Yale University and situates the method in the context of the demise of the classical paradigm, or "golden age," of social psychology. Chapter 2 (Article 1), "Replicating Milgram" (published in the *Journal of Social Psychology* under the title "Revisiting Milgram's cyranoid method: Experimenting with hybrid human agents"), examines the cyranic illusion through replications of two of Milgram's original pilot studies and discusses the method's potential as a means of conducting person perception. Chapter 3 (Article 2), "Echoborgs: Cyranoids with computer program sources" (published in *Frontiers in Psychology* under the title "A truly human interface: Interacting face-to-face with someone whose words are determined by a computer program"), expands upon the traditional cyranoid method by exploring situations wherein a conversational agent (a computer program designed to mimic a human interlocutor) sources for a human shadower, thereby producing a special type of cyranoid known as an "echoborg"; the article places the echoborg within the context of android science, a field that uses humanlike machines as stimuli in social psychological research in order to explore various aspects of human interaction (Ishiguro & Nishio, 2007). Chapter 4 (Article 3), "Using echoborgs to assess intersubjective effort in human-agent dialog" (accepted for publication pending minor revisions in *Computers in Human Behavior*),

combines conversation analysis techniques (e.g., Schegloff, 1992, 1993) with the echoborg method to investigate factors that influence how people repair misunderstandings that arise during dialog with conversational agents. Chapter 5 (Article 4), “Cyranoids in first-person, self-experimental research” (published in *Integrative Psychological and Behavioral Science* under the title “The researcher as experimental subject: Using self-experimentation to access experiences, understand social phenomena, and stimulate reflexivity”), explores the history of researcher-as-subject self-experimentation in social psychology and illustrates how the cyranoid method can be used as a first-person means of directly experiencing the consequences of a transformed social identity through systematic self-experimentation. Finally, Chapter 6, “Cyranoid ethics,” discusses the various ethical concerns involved in cyranoid research, outlines how they were mitigated in the current thesis, and offers suggestions for ensuring positive research participant experience.

As Milgram died before publishing any work on the cyranoid method, and as speech shadowing has seen relatively little application in social psychological experimentation, this thesis attempts to provide the initial basis for future iterations and variants of the method.

ACKNOWLEDGEMENTS

I wish to thank Alex Gillespie, whose guidance over the last four years has been immeasurably valuable. There are few academics within our discipline who would have agreed to supervise a thesis such as this one, let alone pour as many hours into the supervisory relationship as Alex did in my case. Alex's contagious curiosity and penchant for philosophical conversation both broadened and sharpened my thinking immensely. For these things I will forever be grateful.

I want to thank Tom Reader for his supervision and for giving me opportunities to work on research outside of my thesis area over the course of my time at the LSE. These experiences broadened my skills and showed me how research can create impact beyond the laboratory.

I have enormous gratitude for all of my fellow PhD candidates in the Department of Social Psychology, whose friendship and intellectual companionship will last long past our days as students. And to all of my friends and colleagues in the Department of Methodology – thank you. I thoroughly enjoyed teaching alongside you.

I would like to thank the faculty members and administrators in the Department of Social Psychology. I will always try to carry with me your passion for wanting to understand the world as people actually experience it.

I am indebted to those colleagues who assisted me on this research, particularly Geetha Reddy, who spent so much of her valuable time helping to make my experiments possible. I also want to thank my IPOC colleagues – Mark, Meghan, and Brett.

I want to thank Stanley Milgram, whose unfinished work inspired this thesis. Hopefully the research presented herein does justice to the ideas he held so much hope for.

I wish to thank my family, particularly my mom and dad – Vince and Tanice – for supporting me on my academic journey and for trusting my decision to pursue such a path.

Finally, I wish to thank my wonderful Alisa. Your love, patience, kindness, and support sustained me through all the ups and downs of life as a PhD candidate.

TABLE OF CONTENTS

INTRODUCTION	11
CHAPTER 1 – Introducing and Situating the Cyranoid Method	18
A cyranoid is...	18
The origins of the cyranoid method: Stanley Milgram	20
Evaluating Milgram’s cyranoid research	33
Examining the dormancy of the cyranoid method: The passing of the “classical paradigm” of social psychology	36
The argument for revisiting and reapplying the cyranoid method	41
Moving forward	47
CHAPTER 2 – Replicating Milgram	48
Preface	48
Article 1 - Revisiting Milgram’s cyranoid method: Experimenting with hybrid human agents	51
Introduction	51
Background	53
Methodological and theoretical implications of the cyranoid	55
Overview of studies	57
Study 1: Exploring the cyranic illusion in dyadic interactions	59
Study 2: Exploring interpersonal biases with incongruent cyranoids	62
General discussion	76
CHAPTER 3 – Echoborgs: Cyranoids With Computer Program Sources	83
Preface	83
Article 2 - A truly human interface: Interacting face-to-face with someone whose words are determined by a computer program	86
Introduction	87
Contemporary android science	87

The echoborg	94
Overview of studies	98
Study 1: Turing Testing with echoborgs	99
Study 2: A human imitating a chat bot?	103
Study 3: Can covert echoborgs pass as human in the everyday contextual frame?	108
General discussion	113
CHAPTER 4 – Using echoborgs to assess intersubjective effort in human-agent dialog	120
Preface	120
Article 3 - Co-constructing intersubjectivity with artificial conversational agents: People are more likely to initiate repairs of misunderstandings with interlocutors represented as human	121
Introduction	121
Intersubjectivity and intersubjective effort	123
Experimental Study	131
A closer look at the co-construction of intersubjectivity	137
General discussion	147
CHAPTER 5 – Cyranoids in First-Person, Self-Experimental Research	153
Preface	153
Article 4 - The researcher as experimental subject: Using self-experimentation to access experiences, understand social phenomena, and stimulate reflexivity	156
Introduction	156
A brief history of psychological self-experimentation	159
Contemporary researcher-as-subject approaches	162
Milgram's cyranoid method	164
Reflections from researchers who have self-experimented with cyranoids	168
General discussion	175

CHAPTER 6 – Cyranoid Ethics	180
APA and BPS stance on the use of deception in psychological research	180
The nature of the deception used in this thesis	181
How deception was managed: Design and debriefing	182
Participants' reactions to the deception	186
Ethics considerations for future cyranoid research	186
CONCLUSION	193
Limitations of the current work	203
Going forward: Ideas for future work	207
Final remarks	210
REFERENCES	213
APPENDICES	248
A: Experimental materials from Chapter 2	248
B: Experimental materials from Chapter 3	251
C: Experimental materials from Chapter 4	258
D: Example transcripts from Chapter 2	260
E: Example transcripts from Chapter 3	309
F: Example transcripts from Chapter 4	316

TABLE OF FIGURES

CHAPTER 1 – Introducing and Situating the Cyranoid Method

Illustration of a basic cyranoid interaction	19
--	----

CHAPTER 3 – Echoborgs: Cyranoids With Computer Program Sources

Illustration of a basic cyranoid interaction	96
Illustration of a Turing Test scenario involving speech shadowing	102
Illustration of interaction scenarios in Study 2 and Study 3	106
Basic tools of android science	115

TABLE OF TABLES

CHAPTER 2 – Replicating Milgram

Multilevel model fixed factor effect size estimates for utterance length: Study 2	68
Multilevel model fixed factor effect size estimates for question difficulty: Study 2	69
Multilevel model fixed factor effect size estimates for response sophistication: Study 2	69
Comparison of interlocution measures: Study 2	70
Person perception score comparison: Study 2: {[Adult]Adult} vs. {[Adult]Child}	72
Person perception score comparison: Study 2: {[Child]Child} vs. {[Child]Adult}	75

CHAPTER 4 – Using echoborgs to assess intersubjective effort in human-agent dialog

Interactant responses to interlocutor non-repair	147
--	-----

CHAPTER 6 – Cyranoid Ethics

Participants' thoughts on the ethical nature of the study (Chapter 4: "covert echoborg" condition)	187
--	-----

INTRODUCTION

Background and motivation

In the fall of 2011, a little over a month into my M.Sc. program at the London School of Economics, I attended a Department of Social Psychology student-faculty retreat at Cumberland Lodge on the outskirts of London near Windsor Castle. On the first night of the retreat, Dr. Alex Gillespie, a senior lecturer in his first term with the department, gave a presentation on an experiential installation art project he and Robb Mitchell had produced several years prior in Glasgow (Mitchell, 2009; Mitchell, Gillespie, & O'Neill, 2010). The premise of the installation was that of experiencing social life through another body, and of encountering friends, family, and strangers who had themselves been given different bodies through which to interact. The title of the presentation was “Cyranoids.”

A cyranoid, I would learn, is a hybrid social agent composed of the “mind” (or more precisely, the words) of one individual and the body of another, a real-world fusion of separate personae capable of socially engaging with people in all manner of situations. The concept works like this: one person (the “shadower”) wears a discreet inner-ear monitor and uses an audio-vocal technique known as speech shadowing to articulate words transmitted to them by another other person (the “source”). A well-rehearsed speech shadower can perform this technique more or less effortlessly, giving rise to the illusion that they are speaking self-authored words. The shadower thus becomes an interface through which the source interacts with interlocutors (“interactants”) and through which these interlocutors interact with the source. Cyrano had been the brainchild of Stanley Milgram, the social psychologist made famous by his obedience to authority “shock” experiments of the 1960s (Milgram, 1963). Milgram coined the term “cyranoid” in reference to the character Cyrano from the Edmund Rostand play *Cyrano de Bergerac*, which tells the story of a poet (Cyrano) who vicariously fulfills his unrequited love for Roxane by providing to Christian (an inarticulate but handsome nobleman) words with which to woo her.

I instantly became fascinated with the concept of cyranoids. I had come to the LSE seeking to research how one’s external identity impacted their ability to exert social influence, and Milgram’s method immediately struck me as an extremely powerful means of mixing and matching the outer (physical) and inner (dispositional) identities of separate people in order to explore how these different elements affect social perception, first-person experience, and

other social psychological phenomena. What I was drawn to in particular was the fact that the cyranoid seemed to present a means of studying these phenomena in an interactive, highly experiential way. Contemporary experimental methods used to study, for example, how aspects of outer identity affects person perception, are generally quite static and often involve research participants engaging with the *implied presence* of other people via written vignettes rather than involve the *actual presence* of people (Blascovich et al., 2002; Zebrowitz, 2002; Zimbardo, 1999). Where contemporary methods are more interactive, they tend to involve distancing people through computer interfaces or having participants interact in scripted environments so as to achieve a desired level of experimental control. In my estimation, the cyranoid method presented an opportunity to create mundanely realistic yet controllable social experiments given that it allowed research participants to interact in person with a stimulus person (a cyranoid) who had all the attributes of a real person (a non-cyranoid), but whose component parts (the source and the shadower) could be manipulated independently of one another in order to answer a research question about the separate effects of these components.

It came as a surprise to learn following Dr. Gillespie's presentation that Milgram had never formally published any of his work with cyranoids, and the method has remained relatively dormant within psychology since his death in 1984. In fact, Milgram's only concerted effort to publicize his cyranoid studies came in the form of a speech he recorded for an American Psychological Association conference in Toronto a few months before his death, a transcription of which appears in the edited volume of collected Milgram essays *The Individual in a Social World* (Milgram, 2010a). I thus approached Dr. Gillespie with the idea of conducting a Ph.D. that would attempt revive the cyranoid method in order to see what it might offer to psychological research, and this thesis is the culmination of four years of attempting to do precisely that.

Thesis format

The current work has been prepared as a thesis-by-publication in accordance with guidelines established by the London School of Economics' Department of Social Psychology. In addition to this introduction, this thesis includes four academic journal articles (Chapters 2 – 5), two additional chapters (Chapter 1 and Chapter 6), a conclusion, and appendices. Though the use of the cyranoid method connects the journal articles to one another, each article contributes to a distinct literature within social psychology. To enhance continuity and to

make the linkages between the articles more clear, each journal article chapter includes a preface that explains the relationship of the subsequent journal article to the aims of the thesis. I served as the principal investigator, data analyst, and primary author for each journal article, with Dr. Alex Gillespie serving as a secondary author and research supervisor. Two additional secondary authors (Geetha Reddy and Ellen Choi) contributed to Chapter 5's journal article. Each preface includes details as to the proportional contribution of each author. All other sections of this thesis were authored solely by me.

Overview of aims and chapters

Instead of there being overarching theoretical aims to this thesis, the cyranoid method itself will serve as the tissue that connects each chapter as different empirical, historical, ontological, and ethical issues relevant to social psychology are explored. This thesis seeks to lay the groundwork for future research involving the method by demonstrating the unique affordances it can bring to bear on different topics within social psychology, affordances which include allowing for mundanely realistic yet controllable and manipulable person stimuli. This primary aim will be approached via attempting to resolve six sub-aims particular to the thesis's six chapters as well as tertiary aims particular to the various studies contained within each chapter.

The first sub-aim is to present a comprehensive introduction to the cyranoid method and situate it historically. This aim is primarily covered in Chapter 1, "Introducing and situating the cyranoid method," though the method's history is touched upon throughout the thesis. Chapter 1 begins with a description of what a cyranoid is and how one operates within a social interaction via speech shadowing. Due to both the lack of secondary sources detailing the history of the cyranoid method and Milgram's failure to publish details on his cyranoid studies during his lifetime, a thorough accounting of the method's early history required revisiting primary source documents and multimedia contained in Milgram's personal archive at Yale University. Many of these documents have never before been analyzed, let alone woven into the story of Milgram's famous research career. Part and parcel to the history of the cyranoid method has been its dormancy within experimental social psychology since Milgram's death. Chapter 1 considers various reasons for this dormancy, including the paradigmatic changes within North American experimental social psychology that coincided with the end of Milgram's research career and life (i.e., the demise of the so-called classical paradigm, or "golden age," of experimentation). Chapter 1 concludes by exploring reasons

why the cyranoid method is worth revisiting and discusses analogous modern methodologies within contemporary social psychology that might serve as a framework for future cyranoid research.

The second sub-aim comprises validating the cyranic illusion (the phenomenon described by Milgram as failing to perceive when one is interacting with a covert cyranoid; Milgram, 2010a). This is primarily covered in Chapter 2, as well as to a certain extent in Chapter 3. Chapter 2, “Replicating Milgram” (published in the *Journal of Social Psychology* under the title “Revisiting Milgram’s cyranoid method: Experimenting with hybrid human agents”; Corti & Gillespie, 2015b), is an empirical article that includes experimental replications of two of Milgram’s original cyranoid pilot studies. It is demonstrated that even when a cyranoid is composed of highly incongruent component parts (i.e., a child shadower paired to an adult source, and vice versa), naïve interactants reliably fail to notice when their interlocutor is simply speech shadowing. The cyranic illusion provides the cyranoid method with perhaps its greatest affordance: the ability to present multiple and in some cases vastly incongruent individuals together as a singular social actor (a cyranoid) while preserving the interactant’s impression that they are interacting with an autonomously communicating person. Insofar as it is a researcher’s goal to keep a source hidden from an interactant, the cyranoid method can only ever be useful under conditions where participants reliably succumb to the cyranic illusion, and this fact necessitates exploring the robustness of the illusion under various source-shadower pairings. Chapter 2 also examines how interactant addressivity is affected by the composition of the cyranoid they encounter: people adjust what they say to a cyranoid on the basis of the outer identity of the cyranoid. Furthermore, the article demonstrates how interactants anchor onto aspects of outer identity when describing the character of a cyranoid post-interaction and prior to their becoming aware of the true nature of the cyranoid.

The third sub-aim of this thesis is to extend the concept of cyranoids to include scenarios wherein the source of a cyranoid is not a real person, but an artificial conversational agent computer program (a software program designed to mimic a human interlocutor). This special type of cyranoid is referred to as an “echoborg.” The project to develop echoborgs derived from a consideration of whether the cyranic illusion would prove robust in scenarios involving extreme source-shadower incongruity and intrigue as to whether echoborgs might enable humans to interact with machine intelligence through actual human bodies. The main

proofs-of-concept for this derivative of the cyranoid method are described in Chapter 3, “Echoborgs: Cyranoids with computer program sources” (published in *Frontiers in Psychology* under the title “A truly human interface: Interacting face-to-face with someone whose words are determined by a computer program”; Corti & Gillespie, 2015a). Chapter 3 seeks to demonstrate how echoborgs can serve as useful experimental tools within the field of android science, a discipline that seeks to use machine imitations of humans to better understand human social psychology and cognition (MacDorman & Ishiguro, 2006b; Ishiguro & Nishio, 2007). The chapter’s three studies investigate the extent to which echoborgs “pass” as fully human (in the social psychological and sociological sense of identity “passing”; see Goffman, 1963; Renfrow, 2004; Khanna & Johnson, 2010). The first experiment demonstrates how echoborgs can be used in a classic Turing Test (Turing, 1950) scenario. The second experiment explores whether people see echoborgs as autonomous persons speaking self-authored thoughts even when primed to suspect the possibility of encountering an echoborg. A final experiment investigates whether participants take covert echoborgs to be autonomous persons and explores how people make sense of the behaviors of covert echoborgs.

The fourth sub-aim is to show how echoborgs can be used to explore human-agent intersubjectivity, and in particular, how beliefs about an interlocutor’s communicative agency and the interface through which they communicate combine to affect the behavioral patterns exhibited by people during their interactions with various agents. This sub-aim is tackled in Chapter 4, “Using echoborgs to assess intersubjective effort in human-agent dialog” (accepted for publication pending minor revisions in *Computers in Human Behavior* under the title “Co-constructing intersubjectivity with artificial conversational agents: People are more likely to initiate repairs of misunderstandings with interlocutors represented as human”). The chapter presents an analysis of an experiment that uses echoborgs to determine whether the intersubjective effort a person exerts while conversing with a conversational agent (i.e., the scale at which they attempt to repair misunderstandings so as to cultivate common ground) is affected by how the agent is embodied (computer screen vs. human body) and the framing of the agent’s identity (i.e., whether the person explicitly knows they are encountering words determined by an agent). Certain repair activity speech acts (a concept central to Conversation Analysis; Schegloff, 2007), namely *other-initiations of repair* and *self-repairs*, are operationalized as quantitative dependent measures in accordance with a framework recommended by Schegloff (1993), making this experiment one of the few

within the Conversation Analysis tradition to use formal statistical methodology to analyze how conversational phenomena are affected by different experimental manipulations. Furthermore, the article makes a contribution to the literature that combines intersubjectivity theory and human-agent interaction (e.g., Cassell & Tartaro, 2007) by showing how the echoborg method can be used as a means of benchmarking intersubjectivity in human-agent interaction.

The fifth sub-aim is to argue that the cyranoid method can serve not only as a means of conducting traditional, third-person observational experiments, but also as a means of conducting first-person, self-experimental research in which the researcher participates in a cyranoid experiment as either the source or the shadower. This argument is put forward in Chapter 5, “Cyranoids in first-person, self-experimental research” (published in *Integrative Psychological and Behavioral Science* under the title “The researcher as experimental subject: Using self-experimentation to access experiences, understand social phenomena, and stimulate reflexivity”; Corti, Reddy, Choi, & Gillespie, 2015). Whereas the preceding chapters largely focus on understanding interactants’ behavioral reactions to and perceptions of cyranoids, this chapter turns attention to the fact the experiences of the source and shadower can also serve as objects of inquiry. Chapter 5 begins with a discussion of self-observation/self-experimentation in the early schools of psychology (e.g., Wilhelm Wundt’s Leipzig model of experimentation) and how analogues to these models were preserved through the mid-20th century among a cohort of social psychologists that included Milgram and other ethnomethodologists interested in first-person “breaching” experiments (e.g., Garfinkel, 1967). It is argued that first-person self-experimentation can complement contemporary mainstream third-person experimentation and that acting as a source or shadower within a cyranoid in the context of a researcher-as-subject self-experiment affords researchers three potential benefits: (1) access to “social qualia” (i.e., direct experience of a transformed inner or outer identity), (2) improved mental models of social phenomena on account of systematic experiential knowledge, and (3) an enhanced ability to be reflexive about a given experiment via a more complete understanding of the various perspectives involved. Three first-person reflections from researchers who have self-experimented with cyranoids are provided in support of these claims.

The sixth and final sub-aim is to outline an ethics protocol for the cyranoid method given that deception goes hand-in-hand with the use of covert cyranoids. Chapter 6, “Cyranoid ethics,”

details how various ethical concerns were dealt with in each of the studies contained in this thesis and how participants responded during debriefing when the true nature of the studies were described and the identities of the sources revealed. The chapter gives an overview of the important ethical considerations that should be made when designing and conducting cyranoid research and is meant to serve as a guide for researchers planning on adopting the cyranoid method in their own work.

The thesis concludes with a general discussion of the main findings of this thesis and limitations of the method, and provides an outlook for future research applications of the cyranoid method.

CHAPTER 1 | Introducing and Situating the Cyranoid Method

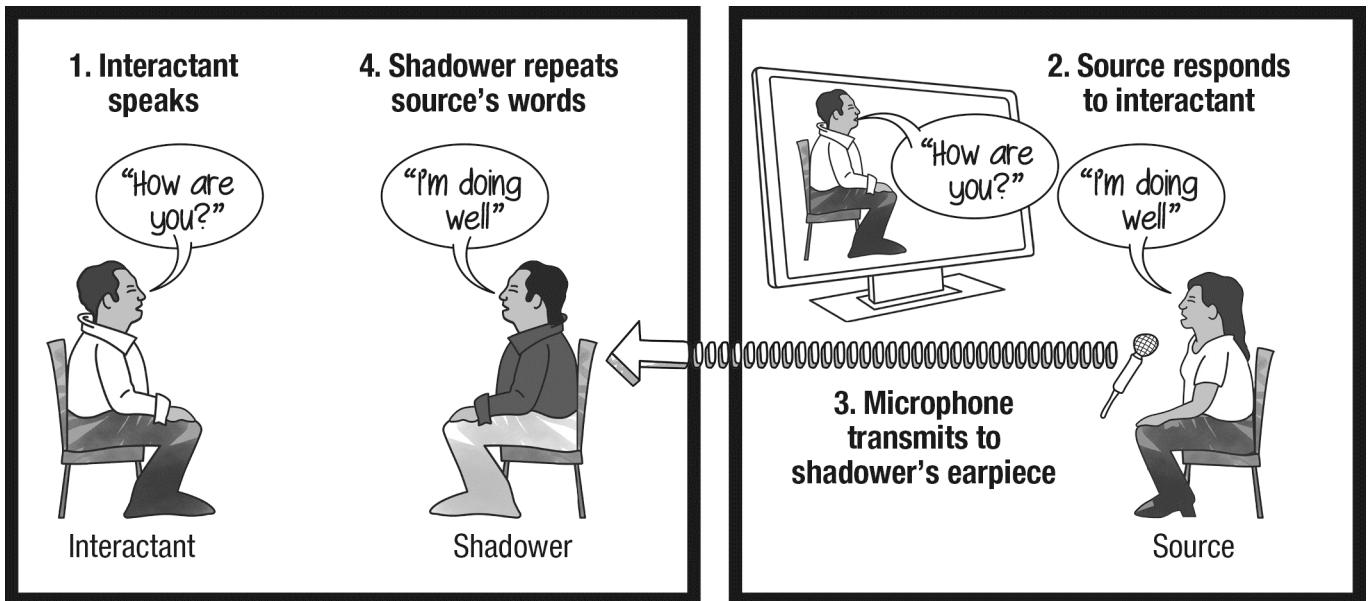
A cyranoid is...

A cyranoid, in its most basic form, is a hybrid social agent composed of two individuals, one of whom behaves as the *shadower*, and the other as the *source*. The source is the metaphorical mind of the cyranoid and supplies the shadower (the body) with verbal communication to instantaneously replicate, and this is accomplished using an amalgam of devices that enable audio produced by the source to be transmitted to an earpiece worn by the shadower. In the studies detailed in this thesis, the shadower remains verbally non-autonomous while ventriloquizing for their source (i.e., they communicate solely the sounds produced by the source and vocalize no self-authored thoughts). Cyranoids engage *interactants* – third parties who encounter the shadower in person but from whom the source may or may not remain hidden. Thus, an elementary cyranic interaction is a tripartite face-to-face social encounter (see Figure 1.1).

To accomplish a fluid cyranic interaction, the shadower must be well rehearsed at speech shadowing – a vocal technique that involves replicating the words one hears in as close to real-time as feasible. The specifics of speech shadowing are discussed in detail throughout this thesis, but the point that will be emphasized here is that shadowing comes quite naturally to native-language speakers. Shadowers can achieve latencies (i.e., the time between the start of a source-utterance and the commencement of replication by the shadower) as low as a few hundred milliseconds (Bailly, 2003; Marslen-Wilson, 1973). This produces an effect whereby from the first-person perspective of interactants, shadowers appear as though they are delivering self-authored prose (a phenomenon known as the “cyranic illusion”). As shall be demonstrated in this thesis, this illusion is extremely robust and can be maintained even in conditions involving high degrees of source-shadower incongruity (e.g., a child sourcing for an adult shadower and vice versa). The means by which a shadower’s earpiece receives verbal stimuli from their source can vary depending on the type of cyranic scenario the researcher wishes to construct; the studies detailed in this thesis involve sources transmitting audio to shadowers by-way-of FM radio transmitters connected to microphones that they speak into.

Functioning as a speech shadower in a cyranic interaction is a dual-task performance requiring both an ability to accurately replicate spontaneous speech received from a source as

Figure 1.1: Illustration of a basic cyranic interaction.



well as an ability to simulate contextually appropriate body language. Though a shadower's body language, if misaligned with the speech they are conveying, may not compromise the cyranic illusion *per se*, misalignment can create a veneer of abnormality that confounds how an interactant perceives the interaction and behaves in relation to the shadower. This notion is underscored by research on the implicit alignment of interlocutors' body movements and speech patterns that occurs during face-to-face interaction, also known as "interactional synchrony" (Bailenson, Yee, Patel, & Beall, 2008; Condon & Ogston, 1966; Kendon, 1970). Any change or disruption to this synchrony may affect the embodied perception of the intersubjective experience on the part of the interactant. Consequently, if a shadower's movements and repertoire of nonverbal cues is to a large degree discrepant from that which is implied by the source's communication, this will in some ways be mirrored by the interactant and may arouse suspicions of non-authenticity. A shadower, thus, must implicitly persuade the interactant as to the genuineness of their communication through nonverbal cues (facial expressions, hand movements, posture, etc.) and avoid gestures that discretely convey subterfuge (for examples see Frank & Ekman, 1997). Taking into consideration one's own perspective of the replicated words is only half of the equation, for the shadower must also instinctively sense which nonverbal behaviors "fit" the intersubjective situation through gaining a sense of what may "feel right" from the perspective of the interactant (see Cesario

& Higgins, 2008).

The task of accomplishing presentation authenticity while shadowing is not as difficult to accomplish as it may sound, however. For a decently practiced shadower, the act of spontaneous speech replication requires very little cognitive load. This is thanks in part to the fact that there is strong evidence for a two-route system of speech production whereby the semantic representation of words is cognitively processed separately from the phonetic and motor requirements of verbal language (McCarthy & Warrington, 1984, 2001). Shadowed speech is facilitated by an input/output loop that bypasses other neural language structures (Bailly, 2003; McLeod & Posner, 1984). As a consequence, shadowers don't necessarily have to spend cognitive resources *thinking* about what to say, and instead may divert attention to convincingly performing nonverbal cues and other motor tasks (see Spence & Read, 2003).

The origins of the cyranoid method: Stanley Milgram

The cyranoid concept originates with Stanley Milgram (2010a), the social psychologist whose obedience to authority “shock” experiments (Milgram, 1963) are among the most widely cited and remain among the most controversial psychological studies ever to have been performed (Reicher, Haslam, & Miller, 2014). During the time he conducted his research with cyranoids (a period which spanned from 1977 until his death in late 1984), Milgram served as Distinguished Professor of Psychology at the City University of New York (CUNY) Graduate Center in Manhattan. He never published material from his work with cyranoids and died a little over a year after he had secured a small research grant from CUNY to explore the ideas he had for a formal set of cyranoid studies. Secondary sources that detail this research are scant and typically offer little more than a basic description of the method. The limited number of more detailed accounts of Milgram’s cyranoid studies have primarily come from Thomas Blass, Professor of Psychology at the University of Maryland. Blass has produced a biography of Milgram’s life (Blass, 2004) as well as edited a well-known collection of Milgram’s which contains a brief chapter on cyranoids derived from a transcription of a speech Milgram recorded for the 1984 American Psychological Association (APA) annual convention (Milgram, 2010a).

Luckily, many of Milgram’s notes, datasets, illustrations, and correspondences concerning his cyranoid studies have been preserved in the Stanley Milgram Papers Archive at Yale University. The portions of this archive that pertain to cyranoids, though highly unorganized,

contain numerous documents that offer a general picture of the arc the method took with Milgram. Though a small component of this content has been briefly described by Blass (2004), this chapter shall offer the first formal survey of the archive's most relevant material.

Beginnings. One of the more substantial documents contained in the archive, and one that provides a glimpse into the origins of the cyranoid method and Milgram's vision for it, is a lab report dated 17 November, 1977 titled "The person as a medium: studies in the cyranic mode" (Milgram, 1977b). In it, Milgram describes having conducted his first cyranic interactions three days prior during his graduate course in mass media. The report begins with a reference to *Cyrano de Bergerac* – the story made famous by playwright Edmund Rostand (1981) in the late 19th century that tells the tale of a poet (Cyrano) who prompts an inarticulate nobleman (Christian) with what to say so that they may jointly woo Roxane. Milgram discusses how he and his students went about constructing a "cyranic contraption" (a term introduced by Mitchell, Gillespie, & O'Neill, 2011) by-way-of an FM radio transmitter that relayed to an inner earphone worn by the "medium" (a term Milgram used early on to refer to the speech shadower). A microphone relayed audio from the interactant to another room where the source was located. According to this document, Milgram's initial explorations with the method involved him and his students taking turns functioning as the shadower, source, and interactant. It is in this document that Milgram first describes the cyranoid method as analogous to mind-body interchange, stating that his and his students' experiences witnessing shadowers was akin to observing a social situation in which "one person is given the mind of another" (Milgram, 1977b, p. 4), and that "it is a somewhat uncanny experience to have a onversation [sic] with someone whose brain is not formulating the words with which he responds to you" (p. 3).

At the time of his initial forays with the method, as this early lab report makes clear, Milgram was still in search of a concrete theoretical domain at which to direct the technique. The curiosity the method roused amongst he and his students lead him to produce a list of potential research directions. An interest in pursuing person perception research topped the list, as Milgram saw the technique as a potentially powerful means of gauging the degrees to which the inner (dispositional) and outer (physical) identities of another person separately influence impression formation. He also suggested that it was worth investigating the experience of the shadower; for instance, to what extent is the shadower aware of what they are repeating? How would functioning as a shadower for a long period of time (days, weeks,

or months) impact the shadower? Furthermore, Milgram was interested in whether or not interactants could detect source-consistency (i.e., if numerous people rotated functioning as the source for a single shadower while in conversation with a single interactant, would the interactant notice the inconsistency? Would it seem as though the shadower had amnesia when sources were rotated?). Milgram was also determined to investigate the experiences of the source, wondering whether or not people who functioned as sources were more or less likely to communicate certain things through the shadower with whom they were paired than through their own bodies.

The list of potential research directions Milgram outlines in this early document is important because it suggests that Milgram was keenly aware from an early stage of one of the great affordances of the cyranoid method: all three components of a cyranic interaction (source, shadower, and interactant) can serve as objects of inquiry for a researcher. Interactants encounter cyranoids, and thus lend themselves to questions concerning social attribution, person perception, stereotyping, and so on. Shadowers experience having an alternative source of thoughts and have to participate in social encounters without verbal agency, thus the cyranoid method presents a novel way of studying phenomena such as cognitive dissonance (e.g., if one is forced to say something their source articulates that is not aligned with their own inner disposition, does their inner disposition subsequently change to match the words spoken?). Finally, from the perspective of the source, one could ask classic questions surrounding phenomena associated with social disinhibition, self-perception, and behavioral confirmation (e.g., does one's behavior as a source align with stereotypes about the shadower?).

Another early document contained amongst the archival material is an extensive transcription of self-narrated audio Milgram prepared sometime between November 1977 and February 1979 (Milgram, n.d.). The document reads very much like a meditation on the cyranoid as a metaphor for the elusive nature of human agency and free will. For instance, Milgram describes a number of real-world social relationships that are cyranoid-like in composition: government (source) ↔ diplomat (shadower), composer (source) ↔ musician (medium), ghostwriter (source) ↔ television host (shadower), and so on. The fundamental phenomena that intrigued Milgram, it seemed, was that of never really knowing the true source of an interlocutor's thoughts, and the reflexive tendency to ascribe authorship of certain thoughts to the person observed speaking said thoughts. He invokes the work of Alan Turing, whose

famous “imitation game” (Turing, 1950) raises the philosophical issue of how one can ever know that they are dealing with a legitimate autonomous person as opposed to a computer should the behavior of computers someday come to perfectly match that of an autonomous person. Milgram draws an analogy to this by wondering whether one could ever truly know that they were dealing with an autonomous person as opposed to a cyranoid should the cyranoid’s behavior be indistinguishable from that of an autonomous person. His growing enthusiasm for the cyranoid method stemmed from the fact that, to him, the method appeared to provide a means of approaching these philosophical questions “with a degree of technical precision” (Milgram, n.d., p. 21).

Despite these metaphysical musings, what becomes clear when reading this transcription is that Milgram was still searching for a suitable scientific domain within which to focus the cyranoid method. Ever the dramatist, some of the hypothetical scenarios involving the technique that he proposes implementing contain more theatrics than psychological substance. For instance, he suggests using the cyranoid method to simulate a hostage negotiation situation wherein a remote expert negotiator would feed lines to a person who, for whatever reason, was a more ideal candidate to physically encounter the hostage taker. He also suggests conducting suicide intervention simulations wherein a speech shadower was used to talk-down a potential victim by channelling the words of someone sufficiently trained to deal with a suicidal person. Milgram believed that these scenarios would test people’s performance in and out of “cyranic mode.” He saw the ability to access and subsequently vocalize the thoughts of a remote expert source as akin to a form of mind augmentation: a person connected to a source, if given the freedom to articulate both their own thoughts in addition to those of their source, would not be constrained by the limited information available within their own central nervous system (i.e., they would have real-time access to a secondary brain, one that was potentially more capable of handling a certain social situation than their own).

Attempts to scale out the method. One of the archive’s key documents is a grant proposal titled “The Technique of Mediated Speech as a Tool in Social Psychology” (Milgram, 1979), which, according to Blass (2004), Milgram submitted in February 1979 to the National Science Foundation’s (NSF) program in Social and Developmental Psychology (the division of the NSF responsible for sponsoring basic mainstream social psychological research). Milgram sought to secure \$200,000 over a two-year period for cyranoid research. The

document opens with what Milgram perhaps intended as a subtle allusion to his Lewinian roots¹, stating that “Certain research methods are useful for the very reason that they alter the ordinary manner in which things work; by doing so they force us into a more rigorous examination of the processes of normal functioning” (Milgram, 1979, p. 1). He follows by directly referencing Garfinkel’s (1964) method of conducting breaching experiments – experiments wherein researchers and confederates directly manipulate a social context by acting in a highly abnormal manner in order to observe how people’s reactions to such breaches expose the web of implicit social norms that are ordinarily taken-for-granted.

In the proposal, Milgram (1979) outlines four research projects he plans to pursue with the grant money. The first project, “Cyranic persuasion,” proposes using the cyranoid method to test whether audiences are more persuaded by a speaker in cyranic mode whose words are determined by a remote expert source than the same speaker functioning autonomously. Study 2, “Application to emergency counselling situations,” suggests using the method within the context of staged hostage negotiations, wherein a non-expert shadower would speak either their own thoughts or those of a remote expert negotiator while being filmed; research participants would then rate whether the shadower is more effective in cyranic mode. Study 3, “Studies in person perception,” proposes using the method to see how the physical identity of the shadower affects how audiences perceive personality attributes of a cyranoid whose source is controlled. Also suggested for Study 3 is an experiment wherein it will be determined whether any discrepancy in personality will be perceived by those who observe a cyranoid whose source changes intermittently (Milgram states that “this question touches deeply on all of human experience. We assume that when we leave a person, then see him again, he is the same person,” p. 9). Finally, Study 4, “The cyranic continuum,” proposes exploring situations of variable shadower-autonomy (i.e., situations wherein the shadower may elect when to interject with their own thoughts rather than strictly repeat solely those of the source they are paired to). In the general discussion of the grant proposal, Milgram conveys that the cyranoid method’s strength is the researcher’s ability to independently vary the source and the shadower in relation to the research question, and at the end of the document he suggests that he wishes to leave open the question of whether or not cyranoids be presented to research participants live or via a pre-recorded video stimulus.

¹ Lewin is credited with suggesting that you cannot understand a system until you try to change it (Schein, 1996).

At best, the grant proposal's four studies each suffer from conceptual shortcomings, and at worst, the experiments Milgram suggests undertaking contain serious design flaws. For example, in describing Study 1 ("Cyranic person"), Milgram fails to convincingly demonstrate why the cyranoid method is needed at all to explore the question of whether or not an expert source's words are more persuasive to an audience when projected through a speech shadower than the shadower's own words. The shadower could, for example, simply articulate two versions of a script (one expert, one non-expert). Furthermore, Milgram failed to notice a glaring confound in this proposed dual-condition experiment: two factors are actually being manipulated in the treatment condition (expert vs. non-expert, cyranic mode vs. non-cyranic mode), not one. The scientific justification Milgram provides for conducting Study 2 ("Application to emergency counselling"), meanwhile, is very weak. He briefly mentions developments in telemetered medical assistance (e.g., when a lay person is able to access experts within a call center in order to attend to a medical crisis) and suggests that shadowers could potentially relay this expert information when dealing with a crisis rather than resort to their own limited knowledge, but Milgram does not offer any reasons why the simulations he proposes would contribute anything of value to the discipline of social psychology. The proposal for Study 4 ("The cyranic continuum"), is very brief and limited to Milgram simply wondering aloud how "effective" a shadower would be in a given situation in relation to the level of input they received from a source (effective in terms of what, however, Milgram does not make clear). The strongest of the grant application's four proposed studies is Study 3 ("Studies in person perception"), as Milgram offers various factorial experimental designs that aim to investigate how differences between inner and outer identity (e.g., a male source paired with female, and vice versa) impact perceived personality – a question with clear social psychological foundations. A significant flaw of Study 3, however, is that Milgram proposes using pre-recorded videos of cyranoids rather than have research participants interact live with a cyranoid. This design aspect undersells the power of the cyranoid method because, again, it could be argued that simply showing participants video-recordings of confederates speaking from scripts is a more controlled means of tackling the same research question.

The archive contains a video Milgram prepared for the NSF grant proposal that features him narrating over various pilot demonstrations of the cyranoid method in his CUNY laboratory (Milgram, 1977a). From one room (the source room) Milgram is shown feeding words to a

male speech shadower in dialog with a female interactant located in an adjacent room (the interaction room). Several minutes into the interaction, Milgram begins swapping different lab assistants in as the source while the camera shifts back and forth between the source room and the interaction room. We witness the interactant continue to engage the shadower while naïve to the fact that the source of the shadower's words has been rotating every few minutes, though it is clear that the interactant has the impression that they are interacting with a coherent, singular individual. In a second video clip, a female speech shadower is used, and this time Milgram reveals to the interactant that their interlocutor (the female shadower) is merely replicating the words of a remote source (an example of an "overt" cyranoid scenario wherein the interactant knows the true nature of their interlocutor). When this is disclosed, the interactant refuses to believe they are interacting with a cyranoid – they have the strong impression that they have engaged an autonomously speaking person. The shadower then reveals to the interactant the radio device that forms part of the cyranic contraption, to which the interactant responds with amusement.

On July 30, 1979, Milgram received a letter from the NSF informing him that his grant application had been rejected. Blass (2004) documents the reasons the various subcommittee members decided to issue a rejection, the common denominator being a lack of theoretical focus. As one reviewer put it:

It seems obvious that the PI [principal investigator] has become fascinated with a method he has invented and is seeking uses to which it can be put. Unfortunately, it seems to me that none of the studies he proposes will illuminate our understanding of social behavior in any substantial way (quoted in Blass, 2004, p 241).

According to Blass (2004), the subcommittee that evaluated the proposal had not watched the videos Milgram prepared, a fact which incensed Milgram. Milgram made minor adjustments to the original proposal and resubmitted it to the NSF in November 1979, this time urging the chair of the NSF subcommittee (Robert A. Baron) to consider the videos prior to making a final determination. Once again, however, the grant proposal was rejected on the grounds that the experiments Milgram suggested would fail to touch upon anything of social psychological significance. In the words of one reviewer, "Milgram is here the clever kid who has been

given a clever hammer and now needs something that needs pounding" (quoted in Blass, 2004, p. 242).

Later years with cyranoids. Despite failing to secure a grant from the NSF in 1979, Milgram continued exploring the cyranoid method through pilot studies with his graduate students. In April 1981, he produced more short videos demonstrating the method, this time using more formal experimental procedures compared to what appeared in his earlier videos (Milgram, 1981). In these newer videos, research participants are shown entering an interaction room where they are told by a research assistant that the study they are involved in concerns how people talk to each other. Participants are asked to engage in a 10-minute discussion with a stranger (a confederate shadower seated across from them) on a topic of their choosing related to current events. In some sessions Milgram acts as the source while in others he acts as the shadower. In several of the iterations showed in the video, once a participant has finished participating in a trial as a naïve interactant, they are instructed by Milgram on how to speech shadow and subsequently perform as the shadower during the next experimental trial. Milgram uses this device to emphasize to the observer just how non-technical a skill shadowing is – people with only a few minutes of rehearsal with the technique are shown to sufficiently perform as a covert shadower while interacting with complete strangers. In one particular trial, the shadower is a local high school student paired with a college student source. Following the interaction, the interactant, an adult naïve to the nature of their interlocutor, remarks that the person with whom they spoke seemed brighter and considerably more articulate than a typical high school student. This particular trial is notable because it reveals the movement Milgram was making toward studying cyranic interactions wherein the source and shadower were significantly incongruent; convinced of the robustness of the cyranic illusion, he seemed to be pushing the method in the direction of creating cyranoids whose inner and outer identities were vastly divergent from one another (in this same spirit, Chapters 2-5 of this thesis involve exploring radically incongruent source-shadower pairs).

The archive contains a number of personal correspondences, notes, and essays on the topic of cyranoids authored by Milgram, his students and his colleagues from around this time. These informal documents help shed light on Milgram's development of the method as well as some of his motivations behind its continued use in his CUNY lab. For instance, in a letter penned to Stuart Albert (a former Assistant Professor of Psychology at CUNY) on October 9, 1983,

Milgram (1983) laments being unclear as to what dependent measure he should focus the cyranoid method upon and wonders aloud whether cyranoids can yield any scientific value, stating: “Sometimes I suspect [cyranoids] more a theatrical than a scientific phenomenon, evoking artistic wonder, rather than serving as a source of scientific propositions” (p. 1). In the same letter, Milgram reveals the connection between cyranoids and his prior work on obedience to authority, stating that in the obedience to authority paradigm, “one person tells another what to do [while in] the cyranoid situation, one person tells another person what to say” (p. 1). Milgram then takes on the critique that the cyranoid method cannot contribute to real world social psychological inquiry given that cyranoids are not real world phenomena; he does so by drawing an analogy to non-Euclidian geometry, “wherein an entire system of logic is worked out based on premises which don’t hold up in the real world” (p. 1-2). Milgram states that such a system can be useful even if its premises don’t hold up in the real world because it forces thinkers to clarify their taken-for-granted assumptions. In his reply (dated November 9, 1983), Albert (1983) outlines a number of suggestions to Milgram and draws a connection between Milgram’s cyranoids and the concepts of self-monitoring and self-perception, remarking: “by monitoring what we say, we may conclude that we are other than we thought, if only because we are formed from the reflections of others, and what is reflected depends in part on what is initially sent” (p. 1).

Another revealing document is a short essay written by Stuart Green (1983) – one of Milgram’s students and lab assistants. Titled “Notes on Cyranoidia,” the essay connects the cyranoid method with Karl Popper’s views on the ontogeny of the self. Popper was known for his distinguishing three domains of real objects (Popper, 1979): World 1 (physical objects), World 2 (subjective psychological states), and World 3 (products of psychological states). Popper’s contention was that in being a product of the human mind, the self is an object that belongs to World 3, yet is embodied in World 1 and is directly experienced in World 2. Green uses this framework to explain the relationship between cyranic experience and self-discovery, and he does so by invoking another of Popper’s precepts, that the self can only be understood so long as it is experienced as in some sense different than the mere body (see Popper & Eccles, 1977). Green (1983) suggests that the cyranoid method provides a proxy for dissociating from one’s material body, and that participating as a cyranoid allows for the experience of “‘self’ and its potencies most fully” (p. 2). Central to Popper’s philosophy is the tenet that human beings have a unique ability to theorize about objects in the world, gather data, and subsequently falsify (or fail to falsify) these theories, and it is

through this process that people develop improved understanding. The self, as an object, undergoes this same scrutiny as it is explored in different contexts and in different embodied formats (e.g., we may discover something about ourselves by observing how we behave when augmenting our bodies or by wearing masks or anonymizing costumes). Green argues that cyranic experiences allow one to directly manipulate the World 1 manifestation of the self, and that such a transformation can allow one to test theories held about the self and thereby become more enlightened as to the self's true nature.

On October 7th, 1982, Milgram applied for a small internal grant from CUNY to formally conduct cyranoid research (a coversheet for the proposal can be found in the archive; Milgram, 1982). Milgram seems to have simply taken the original NSF proposal from three years prior and attached it to his CUNY grant application (albeit requesting only \$11,840) because he suggests conducting the same four studies outlined in the previously-rejected NSF proposal (i.e., emergency negotiation, persuasion, person perception, and cyranic continuum). CUNY awarded him \$5,000 to conduct this research from mid-1983 through mid-1984 (Blass, 2004; Milgram, 1983).

It is unclear whether Milgram ever conducted the studies proposed in the CUNY grant application as the archival materials from this period mostly detail a few small-scale pilot projects involving cyranoids, none of which bear a resemblance to the studies outlined in the proposal. One cyranoid project that appeared to culminate in a formal experiment, however, was a master's thesis exploring the fundamental attribution error conducted by one of Milgram's graduate students, Andrea Martin (n.d.). The thesis, titled "Cyranic attribution: An analysis of attribution error," details an experiment wherein participants were presented with a video stimulus of a person answering a series of questions in either a stereotypically liberal or conservative fashion, with participants either being informed or not informed that the stimulus person was a speech shadower who was simply articulating the thoughts of a source. Participants were asked to rate the stimulus person using a likert-style questionnaire that gauged a liberal vs. conservative political identity. The main hypothesis was that participants would succumb to the fundamental attribution error (i.e., it was predicted that despite the knowledge of a person merely being a mouthpiece for someone else's thoughts, observers would still make attributions toward the person as if the thoughts were their own). The results of the experiment, however, did not show evidence in support of this hypothesis. What is telling about this particular study is that it reveals that even at this point in time, several years

into his forays with the method, Milgram still saw pre-recorded video presentations of cyranoids as a worthwhile experimental stimulus. It is unclear whether Milgram ever encountered the critique that pre-recording negated the necessity of the cyranic technique (again, scripted interactions would more straightforwardly achieve the same effect). The true power of the cyranoid method is that it enables participants to directly interact with a cyranoid in person or have the experience of being a cyranoid (indeed, all explorations of cyranic interactions in this thesis involve in person interaction). Videos of cyranoids, by comparison, are a much weaker stimulus and cause the mind-body discontinuity of the cyranoid to be presented in a significantly more distal way.

Two more substantial pilots exploring person perception that Milgram conducted at around this time, however, did involve research participants interacting in person with a cyranoid. In the first pilot study (Milgram, 2010a), which was designed mainly to gauge whether interactants would succumb to the cyranic illusion, Milgram sourced for four separate shadowers in communication with 20 adult interactants naïve to the deception (with each participant speaking with one shadower). The shadowers Milgram used were a 16-year-old African-American male high school student, a 16-year-old male high school student of Korean descent, a 22-year-old female graduate student (no ethnicity reported), and a 32-year-old Caucasian male graduate student. Interactants were instructed to engage in conversation with their partner on personal and political issues and were provided with a set of questions to guide the interaction. Therefore, although far from scripted, the conversations (which lasted upwards of 20 minutes) did stay within a certain defined range of topics. Following these dialogues, and before disclosing the true nature of the interaction to interactants, Milgram elicited interactants' general impressions of the encounters via a questionnaire composed of a mixture of written-response and multiple-choice items. Milgram only reported select excerpts from these responses, conveying the general sense that interactants were totally unaware of the illusion. The written responses Milgram highlights do indeed suggest that interactants genuinely believed they were speaking to an autonomously functioning individual.

In the second pilot (Milgram, 2010a), Milgram sourced for 11- and 12-year-old child shadowers (recruited from an acting school) whilst being interviewed by panels consisting of as many as six high school teachers naïve to any deception. The teachers were asked to assess their interviewee's intelligence, so in effect were unknowingly evaluating a child with the mind of a professor, so to speak. The teachers were given freedom to generate their own

questions, but were told that they should touch upon areas related to science, literature, and current events. Each panel interviewed both children one at a time, with one of the children in each set speaking autonomously while the other shadowed for Milgram. The children alternated shadowing and speaking autonomously such that each child was interviewed an equal number of times autonomously and as a cyranoid in order to allow baseline condition evaluations to be compared against cyranoid condition evaluations. Following the encounters, Milgram measured the cyranic illusion by prompting teachers with a questionnaire item that read, “I can now tell you that there was a very unusual feature in the interview situations you participated in today. Could you indicate what it was?” (Milgram, 2010a, p. 407). None of the teachers, according to Milgram, came close to identifying the true deception. One teacher, having noticed the earpiece worn by one of the child shadowers, suspected that their interviewee might have had a hearing disability, but did not attribute anything further to the fact that interviewee wore such a device. Milgram concluded that the impression of an authentic personality could be achieved via the cyranic method even in instances when a cyranoid is a “mash-up” of highly disparate individuals.

Aside from the accomplishment of showing that the cyranic illusion holds in such incongruent cases, Milgram’s second pilot highlighted the cyranoid method’s ability to reveal two major social phenomena: (1) how people engage with others on account of the implicit attributions they make (e.g., through acts of addressivity), and (2) how this in turn affects post-interaction judgement. To illustrate, Milgram felt that while sourcing for the young children he was unable to fully demonstrate his intellect due to the fact that the teachers simply didn’t think to ask the types of questions to the children they might otherwise ask to an adult, remarking: “As the source, I was hoping [the teachers] could ask the cyranoid about Freud, Jung, Adler, or at least Darwin and Wittgenstein, but some teachers stuck to fractions and parts of speech” (Milgram, 2010a, p. 407). Due to the low-level questions he received on account of teachers’ beliefs that they were really speaking to a child, Milgram was only able to achieve in certain cases modest increases to the children’s perceived baseline intellect:

When teachers were asked to assign each boy to an appropriate grade level in the fields of English, science, math, art, and social studies, Jason as a cyranoid was elevated at least 2 years in each subject over his autonomous performance and 4 years over his actual grade level, i.e., the sixth grade. In the field of

art, Jason received a mean score of twelfth grade, compared to an eighth-grade level when he was on his own. In social studies, he was elevated from ninth to tenth grade, a gratifying improvement to the source, who, as I have indicated, was a professor of psychology (Milgram, 2010a, p. 406).

Milgram made note of how the social and interactive capacities of the cyranic method allowed for participants to experience embodied, emotional, and deeply personal responses. In their post-interaction evaluations of the cyranoid they believed to be an autonomously communicating child, teachers described feeling embarrassed for not being as intellectually gifted as the child cyranoid, while others spoke of the warmth and sincerity of the cyranoid's personality. One teacher remarked of a child cyranoid they interviewed, "He is sincere in his beliefs but I feel he cannot be opposed graciously" (Milgram, 2010a, p. 407), while another wrote "He was poised and expressive (hands, face). He was fascinating and again for me a little frightening. I feel his self-confidence is almost overwhelming" (p. 406). These types of responses are integral to social experience, yet as Milgram pointed out, too often social psychological methods remove all manner of actual social interaction from the field of study:

We thus see very clearly how the impressions people form are to some extent generated by their own interaction with the stimulus person, the things they bring out and suppress. It is not enough, as some studies in our field do, to present a subject with a prepackaged description of a stimulus person, for in the real world the subject significantly shapes those aspects of the person to which he is exposed (Milgram, 2010a, p. 407).

Milgram's death. Milgram was invited to present his cyranoid research at a symposium titled "New Paradigms in Psychology" in August, 1984 at the APA's annual convention in Toronto (Blass, 2004). This was to be Milgram's first public presentation featuring the method as he had not at that point formally published anything on cyranoids. In July, however, he suffered a heart attack, leaving him too ill to attend the convention. Rather than cancel his presentation, he recorded an audiotape detailing the two person perception pilots reported above and requested that the audio be played to the symposium attendees in lieu of a

live talk. The original audio recording can be found in the Stanley Milgram Papers Archive (Milgram, 1984), a full transcription of which is presented in *The individual in a social world* (Milgram, 2010a). Milgram tried to continue recruiting participants for additional cyranoid studies in the fall semester at CUNY following the APA convention, but from the material available in the archive it does not appear as though anything substantial manifested from these efforts.

On December 20th, 1984, Milgram suffered his fifth heart attack and died at the age of fifty-one. With his death came the end of cyranoid research at CUNY. None of Milgram's doctoral students had used the method in their dissertations (see Blass, 2004, p. 293), and thus none would go on to involve cyranoids in their own research careers.

Evaluating Milgram's cyranoid research

Part of the genius of Milgram, according to Blass (2004), was his ability to invent novel and elaborate forms of experimentation in his pursuit of social psychological questions. This was no doubt the case for his obedience to authority experiment (Milgram, 1963), which entailed both performance theatrics on the part of the confederates employed in the study and the development of an intricate technical apparatus (for a discussion of how Milgram designed his "shock generator", see Oppenheimer, 2013). His "small world" experiments, which sought to explore the number of connections between any two persons in the United States, used a methodology that was highly inventive for the time, involving sending information packets to random recipients in cities in the middle of the country and tracing their pathway back to Harvard University in Boston (Milgram, 1967; Travers & Milgram, 1969). It was in this same creative vein that Milgram developed his cyranoid studies, which combined science with theatrics, artistry, technology, and subterfuge.

From a methodological standpoint, the great accomplishment of the cyranoid pilots was the use of speech shadowing as a social psychological research device in live experimentation. Whereas at the time the technique had been used primarily in psycholinguistics (Marslen-Wilson, 1973) and to a much lesser extent in non-interactive social psychological research (Schwitzgebel & Taylor, 1980), Milgram was the first to report bringing covert speech shadowers face-to-face with subjects to create truly in person cyranic encounters. It is this marriage – speech shadowing with live, in person social interaction – that led to his discovery of the cyranic illusion, the intriguing phenomenon that underpins the utility of cyranoids as a

methodology. Moreover, Milgram's discovery that a child as young as 11-years-old could spontaneously and convincingly speech shadow for an adult demonstrated the power the cyranoid method has for creating hybrid personae from incongruent individuals. Milgram's work with the method also showed that cyranoids can be used for first-person experiential research. As demonstrated by his expressing frustration at not being asked the types of questions that would allow him to display the range of his intellect when interviewed by teachers through the body of children, Milgram (2010a) experienced something of the subjective qualia of having the outward identity of a young child. The cyranoid method is perhaps the first reported means of enabling individuals to experience a face-to-face intersubjective context vicariously through another person's body in real-time.

The major empirical contributions of Milgram's cyranoid pilots are what they add to our understanding of person perception. The fact that the participants in Milgram's studies succumbed to the cyranic illusion so consistently, even when encountering highly incongruent cyranoids, underscores a certain tolerance for ambiguity at work during social encounters with strangers. People simply don't question the communicative authenticity of those they meet unless they enter into the encounter with pre-existing doubts. This reinforces the general social psychological phenomenon that people tend to perceive unity between outer appearance and inner character (Dion, Berscheid, & Walster, 1972; Eagly, Ashmore, Makhijani, & Longo, 1991), suggesting that dichotomies between the mind and body implicitly reconcile in the minds of interactants and observers, particularly during brief casual encounters. In assigning an intellectual score their interviewee, the teachers in Milgram's adult-child cyranoid pilot, while no doubt recognizing the child as extremely intelligent, in most cases scored the child only several years above their baseline (autonomous) scores. Milgram's investigations of the cyranic illusion lend support to the notion of perceptual salience – the principle that describes humans as having a predisposition toward attributing causality to what is most salient to them (Jones & Nisbett, 1972; Taylor & Fiske, 1975).

There are several important shortcomings to Milgram's cyranic work. In particular, it seems as though Milgram never systematically analyzed the majority of data that was collected from the various cyranoid projects he was involved in. He may have gone on to do so had he lived beyond 1984, but unfortunately for both him and the cyranoid paradigm, this was not the case. Aside from the analysis presented in the master's thesis authored by Martin (n. d.), Milgram's use of statistical procedures to compare experimental conditions in his various

pilots seems to have proceeded in a very disorganized and ad-hoc fashion. Milgram collected *a lot* of quantitative data during his time exploring the cyranoid method (as the reams of raw data spreadsheets in the Stanley Milgram Papers Archive will attest), with observations consisting mostly of personality scale responses made by interactants, but unfortunately he seems not to have formally tied this data together in a manner that led to discernible conclusions. Furthermore, despite the dialogical richness produced by the interactive cyranoid pilots, Milgram did not seem to perform even so much as a basic qualitative analysis of these interactions (this thesis, on the other hand, uses a mixture of qualitative and quantitative approaches to arrive thick descriptions of the emergent dynamics of cyranic interactions). In his writings and dictations on cyranoids, Milgram presents very little in terms of concrete evidence in support of the inferences he makes and instead appears to reach deductions via a reliance on instinct; Milgram thus reported his cyranoid research more like a science journalist than a professional psychological researcher. Even in his failed grant proposal to the NSF (Milgram, 1979), Milgram attempted to sell the panel members on the method by-way-of a presenting a clever and theatrical narrative rather than elucidate potential contributions to the field of social psychology via a presentation of preliminary data (quantitative or otherwise) drawn from the pilots he had conducted up until that point.

Another important limitation of Milgram's cyranoid portfolio is that, in his seven years with the method, he largely failed to position cyranoids within the broader context of social psychological research methods and thereby distinguish its unique affordances from those of the other mainstream methods being applied to phenomena similar to those he was investigating. The concept of presenting one person's identity through the body of another individual existed as a method in social psychology prior to Milgram's forays with cyranoids (e.g., the use of "physically attractive" vs. "physically unattractive" images of target-people to test for stereotyping and self-stereotyping in social interaction; Snyder, Tanke, & Berscheid, 1977; also see Chaiken, 1979). As his lamenting of the popular technique of presenting participants with a "prepackaged description of a stimulus person" suggests (Milgram, 2010a, p. 407), Milgram clearly knew that a significant affordance of the cyranoid method absent from other standard techniques was that it enabled interactive face-to-face encounters between participants and stimuli. Curiously, however, Milgram consistently proposed designs for cyranoid studies that consisted of mere videos of cyranoids being shown to participants. What makes his decision to advocate for video-based cyranoid stimuli even more strange is the fact that quite a number of Milgram's more famous and successful

research projects involved real interactions between subjects and confederates (e.g., his obedience studies and his breaching experiments; Milgram, 2010c; Milgram, 2010d). This, with the benefit of hindsight, was a critical error of Milgram's, and it is hardly a surprise that the NSF panel members were so underwhelmed by Milgram's presentation of cyranoids.

Examining the dormancy of the cyranoid method: The passing of the “classical paradigm” of social psychology

Upon Milgram's death, the cyranoid method entered a long period of dormancy, with the studies contained in the subsequent chapters of this thesis being the first formal reapplications of the method within a scientific context. This begs the question as to why Milgram's work with cyranoids went ignored and un-replicated within social psychology for so long. Milgram's other works, particularly his obedience to authority experiments, are widely referenced throughout the social sciences (Miller, Collins, & Brief, 1995) and he is among a small and select group of figures from throughout history whose names and works are near-universal in introductory social psychology syllabi (Miller, 2013). Although he never published material on cyranoids, the second edition of *The Individual in a Social World*, released in 1992, contained the aforementioned transcription of Milgram's pre-recorded APA speech (Milgram, 2010a), so it is not as if evidence of his cyranoid work remained buried deep in his Yale archive and in the memories of those students and colleagues who had witnessed the method first-hand. The method's dormancy is even more puzzling considering the fact that Milgram's work with cyranoids came during the emergence of the dual process theories of social influence, namely the “elaboration likelihood model” (Petty & Cacioppo, 1986) and the “heuristic-systematic model” (Chaiken, 1980). These models attempt, in part, to account for the effects of source-cues (e.g. factors such as age, gender, attractiveness, etc.) on attribution formation, person perception and persuasion. The cyranoid method, with its capacity to nimbly construct and control physical and dispositional traits of a stimulus person for the purposes of live, face-to-face experimentation, could have seen wide application as a research tool that contributed to the dual systems paradigm.

A broad explanation for the dormancy of the cyranoid method can perhaps be arrived at if we step back and consider the historical trajectory of experimental social psychology and Milgram's place within it. Milgram's brand of social psychology derived from those of his mid-20th century “classical paradigm” intellectual forbearers, most notably Kurt Lewin, Harold Garfinkel, Gordon Allport, Muzafer Sherif, and Solomon Asch (who Milgram served

as an assistant to while at Harvard). Asch in particular was known for his predilection for experiments that involved research participants being placed in group contexts populated with confederates and observing how variations on the different facets of the group or the group's environment induced changes in overt behavior (Martin, 2015; see Asch's famous social conformity line experiments: Asch, 1951), a design format Milgram readily adopted. This situationist approach to social psychology was known for being primarily concerned with "empirical demonstrations of the power of immediate situations to affect individuals" (Martin, 2015, p. 25), a research perspective that appealed to Milgram's flair for the dramatic and his preference for method over theory. This penchant was on full display in Milgram's obedience to authority experiments (Milgram, 1963), as he set out to uncover through elaborate laboratory staging how "situational factors might induce obedience in individuals in order to reveal the psychological dimensions of the experience of obedience to a malevolent authority" (Lunt, 2009, p. 25), and was echoed by his situationist contemporaries, most notably Phillip Zimbardo, whose infamous "prison experiment" at Stanford University (Haney, Banks, & Zimbardo, 1973) is perhaps the only other classical era social psychological study to rival Milgram's obedience studies in terms of notoriety and controversy. In both Milgram's obedience research and Zimbardo's prison study, method largely preceded theory, as theorizing as to why participants engaged in certain behaviors followed from the practice of first constructing an immersive experimental situation that functioned as an abstraction of some real-world scenario.

Fundamental to Milgram's situationist approach was his proclivity for using the classroom to generate and test research questions. According to social psychologist Harold Takooshian (2000), one of Milgram's former PhD students, "teaching fed [Milgram's] research... so many of Milgram's most intriguing publications were based on ideas born in his classes – the small world problem, the lost letter technique, the familiar stranger ... [and] cyranoids" (p. 16). Milgram encouraged his students to draw from personal experience when developing research questions (Takooshian, 2000), and believed that social psychologists in general should rely on direct experience "as a source of insight" (Milgram, 2010b, p. xviii). On the basis of such insights, Milgram and his students (*à la* Sherif and Asch) would then devise ways of constructing models of the real-world phenomena they experienced and develop methods for re-creating these phenomena under controlled laboratory conditions or in field experiments wherein they themselves often acted as confederates (Harré, 2006; Milgram, 2010b). This approach allowed Milgram to investigate a broad array of social psychological

topics throughout his career, with each new research question bringing with it the need to invent a novel and often theatrical methodology.

Despite Milgram's successful and wide-ranging research resume, the end of both his life and his research career coincided with a substantial paradigm shift in social psychology, one that saw the field largely move away from the experimental approach Milgram was known for. By the mid-1970s, experimental social psychology was in the midst a self-described "crisis" (Kim, 1999). Serious doubts lingered from both outside and inside the discipline as to the relevance of major research areas (Silverman, 1971). Prominent scholars began to suggest that the targeting of overt human behaviors as dependent experimental measures in an effort to construct predictive behavioral models was problematic given that the types of behaviors that interested social psychologists were fundamentally distinct from the behaviors of objects in the "natural" (non-human) world; human social behavior was prone to change over time as a result of "enlightenment effects," and thus, as some argued, social psychology encompassed more of a historical science than a natural science (Gergen, 1973). The crisis made it increasingly clear that social psychologists lacked sound theoretical guidance, needed to address critiques concerning the ethical components of experimental research, and needed to give new thought to research design and data analysis (Adair, 1991; Elms, 1975). The method most criticized was the traditional laboratory experiment, which, it was argued, produced behavioral artifacts by-way-of researcher effects, demand characteristics, and difficult-to-replicate staging (see Orne & Holland, 1968; Rosenthal, 1966).

Social psychology survived this crisis thanks largely to two parallel developments. First, the emergence of the social constructionist movement served to bring social psychology out of the laboratory and into the field, particularly into communities and cultures where traditional psychological research had not been conducted (Gergen, 1985; Moghaddam, 1987). This both quelled anxiety for the need to account for non-laboratory social psychological phenomena and restored institutional and public faith in the discipline's ability to produce socially relevant impact. Second, lab-based experimental social psychology largely converted to the cognitivist paradigm and moved away from a fixation on overt human behavior as both a dependent measure and as a stimulus (Adair, 1991). These changes helped preserve experimental psychology's status among the traditional, natural sciences. Experimenters became more focused on evaluating various aspects of cognition via indirect social simulation, written vignettes, cognitive tasks, questionnaires, and self-report techniques, with

experimental stimuli generally delivered in a non-social way (often via priming or interaction with a computer interface). These new methods minimized the need for potentially confounding laboratory effects (e.g. direct experimenter/confederate interaction with subjects), reduced the cost of conducting large-scale studies, made experiments more replicable and standardized, and enabled a convergence among researchers on the use of certain psychometric techniques and statistical tools.

Milgram's cyranoid studies thus came about at a time when his own field was largely moving away from the very interactive experimental methods cyranoids are so emblematic of, and as a result, today's social psychology bears very little resemblance to that of Milgram's era, both in terms of the research questions pursued and the methodological tools employed. Milgram and his contemporaries were concerned with, in the words of Zimbardo (1999), "real on-line behavior: what people do, how they act, what they report feeling or thinking in response to the particular situation they have encountered... not research that begins and ends with what people *say* they would do in a given situation" (p. 136). Cyranoids were conceived of as a social stimulus that would confront the research participant face-to-face under varying conditions in order to generate a visceral reaction that manifested in observable behaviors. Social psychology's move toward cognitivism made this style of empiricism old-hat, a shift in emphasis that did not sit well with many classic experimentalists:

For the new breed of cognitive social psychologists, the action is not in gross behavior but in subtle predictions, estimations, judgments, and decisions. A one-paragraph typed description of a scene or setting takes the place of the laboriously constructed stage sets of old-time experimentalists. A check mark on a five-point scale fills in for the buttons on a shock generator or the amount of food or water consumed by a subject. It is so clean, concise, and effortless to construct, to analyse, and to publish that I imagine newcomers to our field wonder why anyone would waste so much time and effort with the old-line experimental approach (Zimbardo, 1999, 138).

Epistemological quarrels between the old and new paradigms aside, the use of deception that was commonplace in experimental social psychology under the classical paradigm met with

greater and greater resistance leading up to the crisis and was one of the principal reasons for the field's movement toward a new style of laboratory research (Adair, 1991; Blass, 2004; Elms, 1975). In 1973, in the wake of the controversy sparked by Milgram and Zimbardo, the APA published its "Ethical Principles in the Conduct of Research with Human Participants" (American Psychological Association, 1973), which in addition to mandating informed consent in research, established the standard that the use of deception should only be used if the reasons for its use are thoroughly described to the research participant following the experiment in order to "restore the quality of the relationship with the investigator" (p. 79). The next year, and in reaction to more serious abuses of research subjects in biomedical research (e.g., the Tuskegee Experiment), the United States Congress passed the National Research Act (1974), which created the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The commission produced the Belmont Report, which established bedrock ethical principles for human subjects research: (1) respect for persons (including the right to informed consent), (2) beneficence ("do no harm"), and (3) justice (the fair administration of procedures). In 1975, the United States Department of Health, Education, and Welfare mandated that all human subjects research be reviewed by an institutional review board (IRB) so as to ensure participant wellbeing (Blass, 2004). The vesting of IRBs with the power to regulate research protocols within institutions, as well as the internalization of ethical norms established by the APA, created a selection pressure for experimental methods that shied away from a dependence on deception, and as such, the use of deception in social psychological research declined after the 1970s (Nicks, Korn, & Mainieri, 1997). Cognitivist social psychology was well-positioned to benefit from this selection pressure as its methods are generally less dependent on what one might consider the most troublesome forms of social deception (e.g., those of the Asch, Milgram, and Zimbardo variety).

Milgram's experiments with cyranoids, thus, came on the heels of substantial efforts made by regulating authorities to curb the use of deception and other research practices that carried with them the potential to inflict psychological harm on participants. Many of Milgram's laboratory and field studies involved deception at some level and a lack of informed consent of today's standard, and his obedience studies in particular fomented a tremendous ethics backlash that has continued through today (see Baumrind, 1964; Nicholson, 2011; Perry, 2012). Cyranic interactions – at least of the kind that are the most psychologically interesting, where the interactant assumes they are engaging with an autonomously communicating and

otherwise normal person – require deception (i.e., they necessitate withholding from the interactant the hybrid nature of their interlocutor). If a researcher sought to use cyranoids in social interactions to gauge the extent to which the inner vs. outer identity of a cyranoid influenced some aspect of interactant person perception, disclosing to interactants prior to the interactions that their interlocutor would be a cyranoid would largely invalidate the conclusions of the study. This is not to say that there are no interesting research questions that can be investigated validly using overt cyranoid scenarios (in fact, several experiments described in the later portions of this thesis make use of such scenarios), but the ability to leverage the cyranic illusion so as instil the belief that one is having an unscripted in person encounter with an autonomous person is a (if not *the*) primary affordance of the cyranoid method. Milgram himself never reported any ethical issues arising from his participants' experiences with covert cyranoids and the deception involved. In fact, video evidence from Milgram's pilots suggests that people, when informed that their interlocutor was a cyranoid, were generally quite amused and enthusiastic about the method (see Milgram 1977a, 1981a). Nonetheless, it is easy to imagine covert cyranoid scenarios where the inherent deception does lead to negative participant experiences.

The post-Milgram dormancy of the cyranoid method, therefore, may have as much to do with the confluence of changes that culminated in social psychology's parting with the classical paradigm, particularly those changes that led to the ascent of taboos concerning the use of deception, as with Milgram's personal failure to formally publish with the method and sell his colleagues on the method's empirical potential.

The argument for revisiting and reapplying the cyranoid method

Contemporary methodological issues in social psychology. It has now been over three decades since the end of the classical paradigm within social psychology, and for that matter, Milgram's last cyranoid study. Some contemporary researchers have lamented the methodological status quo of the new paradigm, particularly the manner in which dependent measures are operationalized and the manner in which supposedly "social" person stimuli have evolved. With regard to the nature of dependent measures, much of the concern surrounds the relative absence of interest in human social behavior of the sort that preoccupied classical era situationists such as Milgram, Asch, Sherif, et al. (see Baumeister, Vohs, & Funder, 2007; Lewandowski & Strohmetz, 2009; Patterson, 2008; Zimbardo, 1999). Baumeister, Vohs, & Funder (2007) summarize the current state of affairs in the following

manner:

The fields of social and personality psychology, however, offer a special and discouraging case. Both of these related fields have a mandate to study the important social behaviors that compose the very texture of human life, with personality psychology focusing on individual differences in those behaviors and social psychology exploring situational influences. But personality psychology has long relied heavily on questionnaires in lieu of behavioral observation, a state of affairs that has begun to change only recently and ever so slowly, at that. Even worse, social psychology has actually moved in the opposite direction. At one time focused on direct observations of behaviors that were both fascinating and important - a focus that attracted many researchers to the field in the first place - social psychology has turned in recent years to the study of reaction times and questionnaire responses. These techniques, which promised to help to explain behavior, appear instead to have largely supplanted it. The result is that current research in social and personality psychology pays remarkably little attention to the important things that people do (p. 396).

Questionnaires and self-report techniques are not problems in and of themselves (indeed, Milgram himself used questionnaires in his cyranoid studies and many of the experiments in the current thesis utilize questionnaires and self-report measures); the problem is the disproportionate use of these modes of data-extraction when making generalizations about the on-line, real-world social behaviors they purport to capture. What the cyranoid method can return to social psychology is a way of conducting truly *social* experiments wherein the observable behavior of interactants and shadowers over the course of a cyranic interaction is assessed in relation to the composition of the cyranoid, the identity of the interactant, and the interactive context within which they meet.

Central to the lament regarding the nature of contemporary experimental person stimuli, on

the other hand, has been a concern with the extent to which today's research tends to sacrifice mundane realism in favor of experimental control (Chapter 5 of this thesis provides a historical overview of this phenomena). In an effort to achieve high degrees of control, many of today's standard research protocols forego concerns about creating contexts that mirror life outside the laboratory (DiFonzo, Hantula, & Bordia, 1998), and as a result, person stimuli in experimental person perception research tend to take the form of "paper-and-pencil" descriptions of people, recordings, or still images of people/people-like stimuli (Zebrowitz, 2002). Allport (1985) famously defined social psychology as the "attempt to understand and explain how the thought, feeling, and behavior of individuals are influenced by the actual, imagined, or implied presence of others" (p. 3). Blascovich et al. (2002) refer to this widely-accepted definition when critiquing what they feel is the unfair neglect of mundane realism in modern social psychology, emblemized by the general lack of realistic social and person stimuli, and the problem this raises regarding the ecological validity of mainstream research:

We find it interesting... that social psychologists have blurred Allport's (1985) presence distinctions (i.e., actual, imagined, or implied), at least in terms of the methods and stimuli they use. Many, if not most, social psychologists apparently assume that empirical reference to and experimental manipulations of actual, imagined, or implied human stimuli are essentially equivalent for understanding social psychological processes... Most would agree that simple written vignettes are far less compelling (i.e., many participants will feel as though they are in an experiment) than the more elaborate staged scenarios (p. 103).

What Blascovich et al. (2002) put forward as a way of striking a balance between mundane realism and control in interactionist research is a research tool that shares many similarities to Milgram's cyranoid method: immersive virtual environment technology (IVET). In immersive virtual environments, research participants can both embody and control 3D human avatars as well as encounter avatars controlled by other research participants, confederates, or artificial intelligence in simulated virtual social environments. One's outer identity can be transformed such that the avatar one controls bears little resemblance to one's own physical traits, and the effect of these transformations can be observed in the form of the

behavioral changes they induce in the other research participants. Proponents of IVET argue that the technology allows research participants to encounter stimuli in a highly dynamic and interactive way that, while virtual in nature, is a more realistic *social* experience for the participant than a vignette or a static stimulus. As with cyranoids, the outer and inner identities of the person stimuli in immersive virtual environments can be independently manipulated. Indeed, proponents of IVET have often cited Milgram's cyranoid studies as a conceptual analogue and source of inspiration (e.g., Bailenson, Yee, Blascovich, & Guadagno, 2008). Over the last decade-and-a-half, the IVET paradigm has been extremely productive at investigating phenomena central to social psychology, including aspects of behavioral confirmation (e.g., Yee & Bailenson, 2007, 2009), stereotyping (e.g., Dotsch & Wigboldus, 2008; Hoyt & Blascovich, 2007), self-representation (e.g., Bailenson, Blascovich, & Guadagno, 2008; Hoyt & Bailenson, 2009), obedience to authority (e.g., Dambrun & Vatiné, 2010), and interpersonal attribution (e.g., Hoyt, Aguilar, Kaiser, Blascovich, & Lee, 2007).

In relation to IVET avatars, the cyranoid method, arguably, provides for much more mundanely realistic person stimuli as cyranoids have real, fleshy human bodies and can create the impression of being an autonomous singular individual. Moreover, the IVET model can potentially serve as a guide to a reborn cyranoid paradigm. Whereas Milgram lacked a methodological analogue to help scaffold the cyranoid method, the productivity of which he could have used to sell his colleagues on the upside of cyranoids for social psychological research (and grant awarding bodies, for that matter), the success of the modern IVET paradigm can be pointed to as a reason for revisiting the cyranoid method.

Recent uses of cyranoids in experiential installations and interactive art. Aside from its relation to the IVET paradigm, an additional reason to revisit the cyranoid method within a scientific context comes from the recent use of cyranoids in experiential installations and interactive art. Alex Gillespie has spearheaded the rebirth of cyranoids in these domains, and together with Rob Mitchell (currently Assistant Professor of Social Interaction Design at the University of Southern Denmark) has used cyranoids as tools for exploring fusions of technology with human life. Mitchell (2009), for instance, has in his own work focused on how simple cyranic contraptions can enable human beings to serve as stand-in communicators for non-present persons (e.g., how a cyranoid can be used as a proxy for a colleague who could not be physically present for a workplace meeting). Mitchell has also

explored the possibility of cyranoids as teaching surrogates in classroom environments (e.g., Raudaskoski & Mitchell, 2013). In one such study, remote teachers sourced for student shadowers in classrooms while delivering a lesson, with the role of the shadower rotating amongst the different students (Mitchell, 2010). Each student therefore had the opportunity to both present to the class in the form of a cyranoid as well as learn via the variety of personalities within the peer group. This activity suggested several potential learning benefits. In particular, students seemed enthusiastically engaged with the learning process, and the exercise demonstrated potential in helping scaffold students' teaching and presentation skills.

While the above-described explorations involve overt cyranoids, Gillespie has also provided the impetus for designing experiential installations involving covert cyranoids. One such installation in an art gallery context involved a number of covert and overt cyranoids intermingling with visitors naïve to their presence (Mitchell, 2009; Mitchell, Gillespie, & O'Neill, 2011). Certain attendees were discretely invited to function as a shadower's source for a period of time from a control room, granting these individuals the opportunity to encounter unsuspecting guests (among whom were their friends and family) through the body of a stranger. Not only did Milgram's cyranic illusion very much hold true in this setting, even those who were approached by an overt cyranoid wearing a large hat with cameras and microphones clearly present succumbed to the illusion. It seemed that even when the nature of the interaction was made explicit to them, interactants still found it difficult to accept that they were actually communicating with someone other than the person right in front of them – a testament to the robustness of the cyranic illusion.

In reflecting upon the art gallery studies, Mitchell et al. (2011) make several observations. First, invoking Garfinkel's (1964) belief in the value of exploring social relations in contexts of grossly violated, or "breached," social norms, they argue that the cyranoid method can serve as a means of constructing positive environments that breach research participants' expectations of normality. Second, the authors highlight that the technique allows for individuals to creatively play within existing social relationships. Third, participant sources seemed to enthusiastically approach strangers through the body of the cyranoid, leading the authors to infer that since the shadower absorbs the embodied sentiments (e.g. embarrassment) that might otherwise restrict social interaction, sources feel less restrained in seeking social encounters with strangers by-way-of a proxy (this possibility had been

suggested by Milgram).

In collaboration with Gillespie and Brian O'Neill (a neuropsychologist), conceptual artist Lucy Pawlak has also made use of the cyranoid concept in a number of instillations and films. Her St. Unicorn's Trust (Pawlak, 2009) performance art piece in London, United Kingdom, featured a “multiplied mind” cyranoid scenario, as visitors to an art gallery unknowingly interacted with a covert team of cyranoids masquerading as a tongue-and-cheek cult spontaneously speaking in unison via a single remote source in response to attendees’ impromptu questions and statements. Meanwhile, her film *Inspection House* (Pawlak & del Paso, 2012) explored a unique cyranic theatrical device: individual characters were distinguished by distinct masks, and throughout the film the actors interchanged which character they were playing such that each scene featured a different combination of actor-character pairs. Playing off of the tendency for people to reconcile mind-body discrepancy (discussed above), a sense of character stability was achieved throughout the story because the audience implicitly associated each character with their mask despite the fact that the actor (source) behind the mask constantly varied. At the film’s premier in London, Pawlak (who was in a remote part of Canada at the time) engaged in a question and answer session with the audience by-way-of a cyranoid surrogate.

These artistic demonstrations, while not scientific in the traditional sense, touch upon many questions of significance to social psychology. As mentioned, such scenarios are examples of the breached environments spoken of by Garfinkel (1964), and strike at the heart of many consequential notions underlying the human experience. In particular, they emphasize the role expectation and physical appearance play in mediating our interactions with others. We may think that what we connect with when we encounter an intimate partner, relative, or close friend is some essential, historically-consistent, and ethereal personal quality that exists beyond their physical nature. But in reality, waking up to find your partner had completely changed bodies while their memory and personality were otherwise intact would, in addition to being quite unsettling, serve to accentuate just how much physical continuity underscores our social relationships.

The historical argument: Milgram’s legacy within social psychology. If for no other reason, the cyranoid method is worth revisiting given Milgram’s historical status within social psychology. One can be agnostic toward or even highly dismissive of Milgram’s

contributions to social psychology (some of his contemporaries saw him as a “dilettante who flitted from one newsworthy phenomenon to the next”; Blass, 2004, p. 259), but his impact on the field and other disciplines cannot be understated. Hagggbloom et al.’s (2002) rank order of the 100 most eminent 20th century psychologists, a list that considered all fields of psychology and considered criteria that included journal and textbook citation frequency in addition to other qualitative measures, placed Milgram at no. 46 (no. 1: Skinner; no. 41: Asch). Beyond the massive impact of his obedience studies, his insights have inspired disciplines far afield. Milgram’s small-world study, for instance, has reached well-beyond his own vision, enjoying success as a source of inspiration in applied mathematics, computer science, epidemiology, and other fields long after his death (e.g., Watts & Strogatz, 1998; see Blass, 2004). Milgram is also one of the very few social psychologists whose ideas and lifework have entered into the greater public consciousness, inspiring popular literature (e.g., Modig, 2003), popular music (e.g., Gabriel, 1986), and popular film (e.g., Almereyda, 2015). No analysis of Milgram as a historical figure, therefore, would be complete without revisiting the cyranoid method. And as has been the case with his small-world study, a reborn cyranoid paradigm may give rise to empirical findings that further Milgram’s status not necessarily as a great researcher, but as an indispensable source of ideas.

Moving forward

The cyranoid method was Milgram’s swansong – a paradigm he never gave up on despite its early critics, dedicated a significant portion of his later years to, and made the topic of his final APA address. Milgram laid the groundwork for the method, stumbling upon the cyranic illusion and testing various permutations of the source-shadower-interactant relationships made possible by the method. But in so many ways the method is incomplete and in need of a rigorous re-evaluation. The real test of the method’s significance, both in historical and contemporary terms, ultimately lies in its potential as a research tool capable of tackling issues of interest to *today’s* social psychologists, and in particular, what it can uniquely contribute to the field where other methods fall short. That being the case, the remaining chapters of this thesis pick up where Milgram left off. Though all of the possible applications of the cyranoid method certainly cannot be explored within the limited scope of this thesis, the technique will be used to interactively investigate a wide but interconnected array of social psychological issues. The next chapter starts the outward journey from Milgram with a replication of the major elements of two of his most comprehensive pilot studies.

CHAPTER 2 | Replicating Milgram

Preface

It was decided that the most sensible starting point for reviving the cyranoid method was to attempt basic assessments of the cyranic illusion and the method's potential as a person perception methodology. To do this, two of Milgram's more substantial pilots – those he described in his APA speech (Milgram, 2010a) – were chosen to serve as the basis for formal replications. This chapter presents those replications in the form of two separate studies.

The first study described in this chapter is predominantly an exploration of the cyranic illusion in a context involving a mental-age-congruent source-shadower pairing. In the two studies described in his speech to the APA, Milgram indicated that no interactant who encountered a cyranoid seemed to suspect the true nature of their interlocutor (i.e., they succumbed to the cyranic illusion). He came about this conclusion via four basic modes of assessment: (1) an analysis of the content of interactant-cyranoid dialog, including via direct first-person observation of the illusion through serving as the speech shadower's source in both pilot studies, (2) an analysis of responses to agree/disagree post-interaction questionnaire items designed to elicit interactants' beliefs as to their interlocutors' communicative autonomy and normalcy, (3) a post-interaction questionnaire item prompting the interactant to disclose their beliefs about the encounter and any possible experimental manipulation, and (4) an analysis of participant statements made during debrief interviewing. This framework functioned as the basis for the assessment of the cyranic illusion in this chapter.

The second study described in this chapter is based on Milgram's (2010a) pilot wherein he sourced for child shadowers in interview contexts; it explores mental-age incongruent cyranic interactions. Though the study also investigates the robustness of the cyranic illusion, the study delves into how interactants perceive incongruent cyranoids as well as how they in turn behave (i.e., *address*) a cyranoid in accordance with these perceptions.

Though Milgram was quite convinced as to the robustness of the cyranic illusion as well as the cyranoid method's utility as a means of exploring social phenomena such as person perception, he only informally reported the results and methods of his pilots, leaving much to be desired in terms of supplying a blueprint for exact replication. Enough information is

described in his APA speech, however, to provide a sound basis for recreating the basic elements of the cyranic interactions he explored. Therefore, while this chapter is framed as a replication of Milgram, it should be noted that certain features of the experiments contained in this chapter differ from those executed by Milgram. None of these differences in any way relaxed the burden of proof for a positive demonstration of the cyranic illusion in relation to those set by Milgram, and in most cases were simply slight deviations from Milgram's original designs made for practical purposes. Additionally, the experiments conducted for this chapter contain a variety of experimental controls and additional manipulation conditions not seen in Milgram's work.

Though the first pilot described by Milgram in his APA speech featured less dramatic age incongruences than those of his adult-child cyranoid scenarios, the main aim of the pilot was to assess the cyranic illusion. Milgram's pilot involved four different speech shadowers (two 16-year-old males, a 22-year-old female, and a 32-year-old male), whereas the experiment that comprises the first study of this chapter only involves one shadower (a 26-year-old male), allowing for a more controlled stimulus. Furthermore, the first study of this chapter compares interactants' experiences in cyranic interactions to those of a control condition (non-cyranic; shadower speaks autonomously), whereas Milgram's pilot does not mention any such use of a control group. Another point of difference is that in Milgram's pilot, interactants were given topic guide of discussion-points to help stimulate conversation involving "personal and political issues, mainly nuclear disarmament" (Milgram, 2010a, p. 404); no such topic guide was used in the first study described in this chapter (participants were free to decide on their own a topic for conversation) so as to allay the possibility of interactants suspecting that their interlocutor (the cyranoid) had rehearsed what to say. As a final point of difference, Milgram himself functioned as the cyranoid's source in his pilot, whereas in this chapter's first study a research confederate blind to the research objectives was recruited to function as the source in the experiment's treatment condition.

The differences between this chapter's second study and the second pilot described in Milgram's APA speech consist of the following. Milgram's study involved two child shadowers (an 11-year-old boy and a 12-year-old boy), whereas this chapter's second study featured one 12-year-old boy. While Milgram's study involved two conditions (a treatment condition in which he sourced for child shadowers and a control condition in which the child shadowers spoke autonomously), this chapter's second study involved two additional

conditions: a second treatment condition in which the child sourced for an adult shadower and a second control condition in which the adult shadower spoke autonomously. As a final point of difference, Milgram recruited actual teachers to participate as interviewers in his study, whereas in this chapter's second study interactant interviewers were adults recruited from the greater-London area.

It should be noted that the two studies described in this chapter are highly exploratory in nature. As such, though statistical analyses are performed on several dependent measures, no formal hypothesis testing is carried out. The spirit of these studies was to investigate the cyranoid method as Milgram conducted it, and where appropriate, offer theoretically-informed reflections on the phenomena that emerged; these reflections are outlined in the general discussion section that concludes this chapter. The general discussion section also outlines a number of possibilities for future cyranoid research.

The journal article that follows was authored by Corti and Gillespie (2015b). Corti designed and carried out the studies, performed the data analysis, outlined the article, and authored the main drafts, contributing roughly 75% of the content. Gillespie provided the technological devices, helped conduct the studies, provided key supervisory assistance, provided funding, provided Corti with editorial suggestions for the article, and contributed roughly 25% of the content. An editor at the *Journal of Social Psychology* and two anonymous reviewers provided helpful feedback, suggested edits, and ultimately approved the article for publication.

Copies of participant instructions and other materials used to conduct the studies in this chapter can be found in Appendix A. Example transcripts from the experimental conditions described in this chapter can be found in Appendix D.

CHAPTER 2 | Replicating Milgram

Article Title

Revisiting Milgram's cyranoid method: Experimenting with hybrid human agents

Authors

Kevin Corti (primary)

Alex Gillespie

Abstract

In two studies based on Stanley Milgram's original pilots, we present the first systematic examination of cyranoids as social psychological research tools. A cyranoid is created by cooperatively joining in real-time the body of one person with speech generated by another via covert speech shadowing. The resulting hybrid persona can subsequently interact with third parties face-to-face. We show that naïve interlocutors perceive a cyranoid to be a unified, autonomously communicating person, evidence for a phenomenon Milgram termed the "cyranic illusion." We also show that creating cyranoids composed of contrasting identities (a child speaking adult-generated words and vice versa) can be used to study how stereotyping and person perception are mediated by inner (dispositional) vs. outer (physical) identity. Our results establish the cyranoid method as a unique means of obtaining experimental control over inner and outer identities within social interactions rich in mundane realism.

Keywords

Cyranoid, embodiment, Milgram, mundane realism, person perception, stereotyping

Introduction

In Edmund Rostand's play *Cyrano de Bergerac*, Christian, a handsome yet inarticulate young cadet, woos the love of Roxane by speaking to her the graceful prose of Cyrano, a man whose unremarkable physical features instil in him a paralyzing sense of self-doubt (Rostand, 1981). Through Christian's body, Cyrano achieves a means of vicariously fulfilling his unrequited love for Roxane, while Christian is in turn the beneficiary of ghost-written words that garner affection. This well-known story is but one of the many examples of a fantasy that has appeared in the arts and mythology throughout history - that of the fusion of separate bodies

and minds. Other illustrations include *The Wonderful Wizard of Oz*, in part the tale of a fraudster who is able to attain great power by presenting himself to the world through an intimidating artificial visage. The film *Big* entertains the folly that ensues when an adolescent boy awakens to find himself in the body of a middle-aged man. More recently, films such as *Avatar* and *Surrogates* have imagined hypothetical futures in which mind can be operationally detached from body, allowing individuals to operate outer personae constructed to suit their social goals. Fiction though they may be, these stories illuminate the power façade has over how we are perceived by ourselves and by others, and how we and others in turn behave in accordance with these perceptions.

Stanley Milgram, perhaps best known for his obedience to authority experiments (Milgram, 1974), operationalized the *Cyrano de Bergerac* paradigm in a series of pilot studies conducted shortly before his death. In these pilots, he explored constructing hybrid social agents, whom he called “cyranoids” (in reference to Cyrano), via a vocal technique known as “speech shadowing,” a procedure in which a person immediately repeats auditory stimuli originating elsewhere. Milgram’s idea was to have one person (the “shadower”) replicate the spontaneous speech of another (the “source”) via a covert audio-relay apparatus while socially engaging with research subjects (the “interactants”) naïve to the subterfuge, and his findings suggest that interactants will fail to detect that their interlocutor is a cyranoid. This “cyranic illusion” persisted in cases of extreme identity incongruity between source and shadower, such as when he sourced for child shadowers being interviewed by groups of teachers, none of whom believed following these interactions that they had been talking to anything other than an autonomous (albeit unusually bright) child. Milgram never formally reported the results of these studies, though descriptions of them can be found in a speech he prepared for an American Psychological Association (APA) convention in 1984 (Milgram, 1992) as well as in a biography authored by Blass (2004). In his APA speech, he expressed optimism that the cyranoid method could evolve into a powerful means of researching the social self and person perception. Despite this enthusiasm, no experimental validation of the method has to-date been reported, rendering cyranoids a largely dormant part of Milgram’s legacy.

Our goal in the present work is to resurrect the cyranoid method by exhibiting its utility as a social psychological research tool. In two studies based on Milgram’s original pilots, we examine the robustness of the cyranic illusion and demonstrate how with the method one can

explore various aspects of person perception and the role of stereotypes in social behavior. The aim is to stimulate further research into the wide range of social and cognitive phenomena that lend themselves to investigation by-way-of cyranoids.

Background

Speech shadowing. A functioning cyranoid is a synchronized performance between two or more people and depends upon the shadower reliably and rapidly repeating the words of their source without revealing the true nature of the communication to interactants. This, however, is not as difficult a task as one might suspect, as studies have shown speech shadowing to be a surprisingly simple undertaking. Marslen-Wilson's (1973) early work exploring speech shadowing latencies influenced Milgram's conceptualization of the cyranoid, and the technique has since been used to investigate phenomena ranging from secondary language acquisition (e.g., Murphey, 2001) to speech pathology (e.g., Harbison, Porter, & Tobey, 1989; Healey & Howe, 1987) to cognitive linguistic processing (e.g., Fowler, Brown, Sabadini, & Weihing, 2003). Native language shadowers can track the continuous familiar prose of a source at latencies as low as 70 milliseconds (Bailly, 2003), and continuous unfamiliar prose at latencies as low as 250 milliseconds (Marslen-Wilson, 1985). Shadowers tend to reflexively mimic gestural elements of their source (Fowler et al., 2003; Goldinger, 1998; Mitterer & Ernestus, 2008; Shockley, Sabadini, & Fowler, 2004), while listeners tend to perceive more acoustic-phonetic similarity between persons A and B when A is shadowing for B than when A is speaking non-shadowed speech (Namy, Nygaard, & Sauerteig, 2002; Pardo, Jordan, Mallari, Scanlon, & Lewandowski, 2013), evidence for a phenomenon known as "phonetic convergence." Thus, in addition to replicating pure syntax at low-latency, shadowers instinctively mirror their sources' idiosyncratic speech qualities.

Schwitzgebel and Taylor (1980) explored speech shadowing as a social psychological experimental tool when investigating aspects of third party impression formation. Their shadowers were able to effectively convey both verbal and nonverbal cues necessary for positive impression formation while replicating the words of others. While experimental stimuli in these studies were short videos of shadowers, the authors do report piloting the shadowing procedure *in vivo*. Milgram (1992) referenced Schwitzgebel and Taylor's study as an example of how speech shadowing could be used in social experimentation, but his ambition was to employ the method in interactive settings where research subjects freely dialogued with shadowers face-to-face.

Milgram's pilot studies. In the first pilot described in his APA speech, Milgram (1992) reports having 20 naïve participants engage in one-on-one conversations with various adult cyranoids for whom he sourced, and following these interactions no participant agreed with a questionnaire item suggesting that their interlocutor had been merely repeating messages received via radio. Upon learning the true nature of these interactions, some participants “felt the loss of a person,” having had quite an engaging experience with their interlocutor, who, as it turned out, was merely a “synthetic creation of the experimental procedure and had no existence apart from the hybridization which the experiment created” (Milgram, 1992, p. 340). Notably lacking from this study were control groups (specifically, non-cyranoid dyads) capturing participants’ baseline experiences with the shadowers.

Milgram suspected that interactants would still be inclined to see a cyranoid as autonomous even in cases where a source and shadower were quite dissimilar from one another. Accordingly, he tested the robustness of the cyranic illusion by conducting the aforementioned interview-panel study wherein he separately sourced for 11- and 12-year-old shadowers while being interrogated by groups of teachers. The teachers were asked to assess their interviewee’s intelligence during the interviews, so in effect were unknowingly evaluating a child producing the words of a university professor. Rather than provide a systematic analysis of these interactions, however, Milgram reports select anecdotes from teachers’ post-interview written evaluations highlighting how the deception went undetected despite the conversations being very incongruous.

Cyranoids after Milgram. Despite being largely ignored within the scientific community, the cyranic technique has recently been picked up by artists who have used cyranoids as parts of social installations within which participants experience breaches of social norms (Mitchell, Gillespie, & O’Neill, 2011; Pawlak, 2009) and that create conditions under which people unknowingly encounter familiar others (e.g., friends and spouses) through the bodies of strangers (Mitchell, 2009). The cyranoid has also been used as a metaphorical device within societal and media analysis to describe public perception of highly visible social actors (e.g., the movie star, the news anchor, the politician, etc.), whose relationships with the masses are often mere performance and whose messages are often carefully crafted by unseen speechwriters (McCarthy, 2006, 2011). Despite these developments, which touch upon phenomena fundamental to social psychology, experimental scenarios involving human cyranoids have not yet been formally investigated.

Creating a cyranoid. There are many combinations of gadgetry that might facilitate a functional cyranic interaction. However, a researcher attempting to construct a cyranoid must make considerations based on the level of mobility and covertness they hope to attain. Various technologies enable low-latency audio transfer between source and shadower (e.g., radio transmitters, Wi-Fi, mobile phone devices, etc.), each with certain benefits and drawbacks. Inner-ear radio receivers similar to those used in the current work provide perhaps the greatest degree of mobility and stealth as they are wireless and not readily perceivable at close distances by interactants. Audio relay from shadower to source can easily be accomplished using wireless microphones. Though video relay from shadower to source is not necessary, providing the source with a feed of their shadower’s field of vision gives a richer sense of the intersubjective phenomena occurring between shadower and interactant. Live video relay can be accomplished via overt gadgets (e.g., subcams; see Lahlou, 2011) or covert recorders. Further descriptions of the types of gadgetry that can constitute a “cyranic contraption” are discussed by Mitchell et al. (2011).

Methodological and theoretical implications of the cyranoid

The cyranoid method holds particular promise as a means of constructing and controlling the inner (dispositional, non-visible) and outer (physical, visible) identities of human stimuli in experiments that approximate real-world scenarios (e.g., unscripted, face-to-face interlocution). The methodological advances in social psychology that followed Milgram’s era established experimental norms that prioritized internal validity and replicability generally at the expense of mundane realism and Milgram-esque experimental flair (Adair, 1991), and as such, most modern substantiations of the field’s major theories (e.g., the dual processing models of information processing; Chaiken, 1980; Petty & Cacioppo, 1986) have relied upon static experimental stimuli (e.g., vignettes and cognitive tasks) in which research participants are largely isolated from anything resembling an actual *social* context. Where methods do involve dynamic human-human interaction, potential confounds are often reduced by physically distancing participants from human stimuli (e.g., via computer-mediation) or by restricting dialogue within strict parameters by-way-of role-playing and/or scripting. As social psychological methodologists point out, the de-socializing of experimental stimuli has arisen largely due to the need to control confounds and preserve independence among observations of dependent variables—prerequisites for standard analytical techniques such as ANOVA (see Kashy & Kenny, 2000; Willard, Madon, Guyll, Scherr, & Buller, 2012).

The cyranoid method enables specific forms of experimental control to be introduced into research scenarios involving participants more or less freely associating with human stimuli (cyranoids) face-to-face and in close-proximity. This affords researchers the opportunity to achieve levels of mundane realism not possible with traditional stimuli (e.g., “paper people” and the like; see Murphy, Herr, Lockhart, & Maguire, 1986). A well-trained shadower can spontaneously replicate the prose of a large variety of source-types, constituting a controlled outer identity (or “body”) across experimental conditions differentiated by inner identity. Likewise, a single source can serve as a controlled inner identity (or “mind”) across experimental conditions differentiated by shadower-type. In fact, this logic has recently inspired researchers operating in the overlap between social psychology and computer science to develop an experimental practice that makes use of immersive virtual environment technology (see Bailenson, Beall, Loomis, Blascovich, & Turk, 2004; Bailenson, Yee, Blascovich, & Guadagno, 2008; Blascovich et al., 2002). These researchers argue that this technology offers a means of achieving high levels of mundane realism and experimental control, and often cite Milgram’s cyranoid method as an analogue. In immersive virtual environments, participants control human avatars (the digital equivalents of cyranoids) in three-dimensional simulated social worlds as researchers observe how users’ behaviors and perceptions change in relation to the characteristics of the avatars they ostensibly control and interact with (see “the Proteus effect”: Yee, Bailenson, & Ducheneaut, 2009; “walk a mile in digital shoes”: Yee & Bailenson, 2006).

The cyranoid method also presents a means for social psychologists to examine a number of core theoretical paradigms. One such paradigm, which we consider in Study 2 of the present work, involves the role of appearance cues (e.g., age, gender, height, ethnicity, etc.) in mediating person perception. For instance, it has been well documented that people tend to implicitly perceive unity between outer appearance and inner disposition (e.g., attractiveness ↔ competence: Dion, Berscheid, & Walster, 1972; Eagly, Ashmore, Makhijani, & Longo, 1991; race ↔ aggression: Duncan, 1976; Sagar & Schofield, 1980; youthfulness ↔ naivety; Berry & McArthur, 1986). Using the cyranoid method, one may investigate how interactants’ perceptions of a cyranoid interlocutor change when the identity of the shadower is manipulated (e.g., by age or gender) and the source is kept constant. We can thus come to understand the components of a target source’s disposition that are perceived as more or less stable irrespective of outer identity (i.e., which elements “cut through” the exterior), and those that are susceptible to change according to outer identity.

Connecting these issues with related literature addressing social behavior, there has been well-established research on stereotypes and their often self-fulfilling nature—how a perceiver's biased expectations regarding a target on the basis of their appearance may actually elicit stereotype-confirming patterns of behavior from targets (Snyder & Stukas Jr., 1999; Snyder, Tanke, & Berscheid, 1977) as well as jointly influence behavioral confirmation by perceivers (Chen & Bargh, 1997). These issues can be investigated in highly dynamic and interactive contexts using cyranoids, as the identities of sources, shadowers, and interactants can be manipulated to test, for example, how the intersubjective phenomena that arise between cyranoids and interactants change on account of the dispositional and physical makeup of a cyranoid. Though Milgram had in fact alluded to stereotype phenomena in describing the outcomes of his pilots, he never couched his observations within a broader theoretical framework (Blass, 2004).

Overview of studies

Studies 1 and 2 are modeled off of the pilots Milgram (1992) conducted, though each goes beyond Milgram in terms of scope, control, and breadth of analysis. Both studies investigate the robustness of the cyanic illusion using a number of approaches, including post-interaction interviewing, survey-response, and video/transcript review. Study 2 examines aspects of person perception in relation to inner vs. outer identity and associated behavioral phenomena. In both studies, confederates were trained to function as sources and shadowers across various experimental conditions while participants served as naïve interactants. Each study was separately approved by an ethical review board at a major British university. Studies were conducted in a behavioral research laboratory and participants were recruited from a major metropolitan area via internet advertisement.

The aspect of identity we manipulate in Study 2 is age group (child vs. adult) on the basis of it being the trait dimension explored by Milgram. It has been shown that age is a characteristic with which individuals reflexively categorize others into person-types (Brewer & Lui, 1989), and that people tend to define themselves and others relative to prototypes representative of discrete age groups (e.g., infant, young-adult, middle-aged, etc.; Giles & Reid, 2005). Based on a target's overt age, people instinctively make judgments concerning a variety of socio-personal dimensions, such as social status and competence (see Berry & McArthur, 1986; Brewer, Dull, & Lui, 1981; Fiske, 2010; Krueger, Heckhausen, & Hundertmark, 1995). In particular, people are more likely to ascribe intelligence to a target if

the target shows certain signs of aging (see Montepare & Zebrowitz, 1998; Muscarella & Cunningham, 1996). We use interactions involving cyranoids composed of age-discrepant source-shadower hybrids to observe how interlocution behavior and person perception align with age-based stereotypes.

Cyranoid notation. The terms “source,” “shadower,” and “interactant” have meanings in contexts outside of cyranoid research that may obfuscate their usage herein. Therefore, we have devised a notation scheme for illustrating cyranic interactions that we shall employ in conjunction with the terminology.

There are three essential components to a cyranic interaction: the source (the agent who relays communication to the shadower), the shadower (the agent who shadows speech provided by the source), and the interactant (the agent who physically encounters the shadower). The hybrid persona that results from merging a source’s words with a shadower’s body is called a cyranoid. We use braces (“{}”) to distinguish what is visible to an interactant (namely, the body of either a cyranoid or an autonomously communicating interlocutor), and square brackets (“[]”) to denote the source of the body’s communicated words (which can be either their own if speaking autonomously or that of a third party when shadowing):

{[Joe]Joe}	Joe speaking self-authored words
	{[Joe]Ben} Ben shadowing
	Joe’s words (forming a cyranoid)

General descriptors and subscripts can be used to describe the makeup of an agent:

{[Adult]Adult}	An adult speaking self-authored words
{[Adult]Child}	A child shadowing for an adult
{[Female1]Female2}	Female2 shadowing for Female1

The addition symbol (“+”) can be used to describe multiple sources and multiple shadowers:

$\{[A + B]C\}$ C shadowing for both A and B

$\{[A]B\} + \{[A]C\}$ B and C both shadowing for A

Left-right arrows (“ \longleftrightarrow ”) are used to distinguish communication with interactants:

$\{[A]B\} \longleftrightarrow \text{Group C}$ B shadowing for A in dialogue with Group C

$\{[A]B\} \longleftrightarrow C + D$ B shadowing for A in dialogue with C and D

Finally, general descriptors and subscripts can signify type-similarity/dissimilarity:

$\{[\text{Female}_{\text{USA}}]\text{Male}_{\text{UK}}\}$ Male (British) shadowing for Female (American)

Study 1: Exploring the cyranic illusion in dyadic interactions

The goal of our first study was to validate the cyranic illusion through a simple experiment designed to gauge whether participants would detect a speech shadower during face-to-face, close-proximity, unrehearsed, dyadic interlocution. Aside from Milgram's (1992) pilots, there was no precedent for studying such a phenomenon in the psychological literature, so the techniques we used to investigate the illusion borrowed principally from Milgram.

Method. Subjects and confederates. Forty adults participated in the study (22 female; mean age = 30.25; $SD = 9.95$). Two confederates partook in the study: a 23-year-old female graduate student and a 26-year-old male graduate student. Participants were randomly assigned to experimental conditions. Both confederates engaged in numerous mock trials of the experimental procedure so as to become familiar with the apparatus and achieve a consistent degree of accurate, low-latency speech replication.

Conditions. The experiment featured two conditions: Cyranoid ($n = 20$ participants), and Non-cyranoid ($n = 20$ participants). In the Cyranoid condition, participants individually

engaged in a 10-minute face-to-face dialogue with the male confederate who participants were led to believe was another research participant when in fact the male confederate was speech shadowing for the female confederate for the entirety of the interaction: {[Female]Male} ↔ Participant. The Non-cyranoid condition mirrored the protocols of the Cyranoid condition, albeit this time the male confederate spoke autonomously with participants: {[Male]Male} ↔ Participant.

Instructions and materials. Participants were informed that the study involved holding a 10-minute conversation with another research participant. In order to convey the notion that the interactions were not scripted, participants were told that during the conversation they could speak with their interlocutor about whatever subject they wished. Following these interactions, participants completed a brief questionnaire that included three items designed to gauge their suspicions as to whether or not they believed their interlocutor was (a) simply repeating messages received via radio, (b) giving scripted responses, and (c) speaking their own thoughts. Participants were asked to either “agree” or “disagree” with each statement. Participants were also asked to provide a brief written evaluation of the person with whom they communicated, the purpose of which was to capture subjective, spontaneous impressions of the interlocutor.

Apparatus. The male confederate sat roughly 6 feet apart from participants in an interaction room. Video cameras were placed at eye-level behind each interlocutor so as to capture approximations of their fields of vision. Adjacent to the interaction room was a soundproof source room (used for Cyranoid trials) which housed devices that enabled the female confederate to source for the male confederate. The video camera capturing audio and the field of vision of the male confederate was broadcast to a video monitor in the source room so that the female confederate could observe activity between the male confederate and the participant.

The cyranoid contraption consisted of a series of connected devices. The female source spoke into a microphone connected to an FM transmitter that relayed to a small radio worn in the pocket of the male confederate. Connected to the radio was a neck-loop induction coil worn under the confederate’s clothing. The male confederate wore a discreet, flesh-colored, wireless, inner-ear audio device (not noticeable at close distances) that received an audio signal from the induction coil. This apparatus allowed the male confederate to hear the words

of the female confederate in the source room in real-time.

Procedure. Following informed consent and instruction, the participant was led to the interaction room. The male confederate then entered the room and sat facing the participant. The researcher left the room and participant-confederate dialogue commenced. After 10 minutes, the researcher returned to the interaction room, instructed the male confederate to leave, and delivered the post-interaction questionnaire to the participant. Following completion of the questionnaire and written evaluation, the participant was interviewed by the researcher and asked verbally whether they detected anything unusual regarding their interlocutor's verbal and nonverbal communication. Following all attempts to extract their suspicions, the researcher disclosed the full nature of the study in a debrief session.

Results. Using *STATA*, participants' agree/disagree questionnaire responses were analyzed using exact logistic regression, an alternative to binary logistic regression that provides a more accurate model of small samples (Hirji, Mehta, & Patel, 1987; for procedure see UCLA Statistical Consulting Group, 2014). In relation to the Non-cyranoid condition, participating in the Cyranoid condition did not significantly affect the odds ($\pi_{\text{agree}}/\pi_{\text{disagree}}$) of a participant agreeing with the questionnaire statement: "My interlocutor spoke their own thoughts," OR = 0.31, SE = 0.37, 95% CI [0.01, 4.24] (CI spanning 1 signifies non-significance). Overall, 19 of 20 participants in the Non-cyranoid condition indicated agreement with this item compared to 17 of 20 participants in the Cyranoid condition. Participating in the Cyranoid condition did not significantly affect the odds of agreeing with the statement: "My interlocutor gave scripted responses," OR = 0.31, SE = 0.37, 95% CI [0.01, 4.24]. Three of 20 participants in the Non-Cyranoid condition indicated agreement with this item compared to only one of 20 participants in the Cyranoid condition. No participant in either condition agreed with the questionnaire item that read: "My interlocutor spoke by receiving radio messages and repeating them."

No participant stated that their interlocutor was behaving unusually or in a pre-prescribed manner during post-interaction interviews and debriefing. Moreover, none of the written evaluations provided by participants in the Cyranoid condition gave any indication that the cyranic illusion was detected, and when the deception was revealed during the debrief session, responses were a positive mixture of astonishment and amusement. As further

evidence of the illusion, a review of the video recordings and dialogue transcripts showed that at no point during any of the experimental trials did participants raise the possibility that their interlocutor was talking via assistance or using a script.

Study 2: Exploring interpersonal biases with incongruent cyranoids

Study 2 was modelled off of Milgram's (1992) second pilot and featured two age-discrepant male confederates (a child and an adult) interviewed by panels of participants both autonomously and interchangeably as sources and shadowers for one another. Whereas the sole focus of Study 1 was to demonstrate the cyranic illusion, our second study included the additional goal of examining how verbal behavior confirms age-group stereotypes. To this end, we considered three components of cyranoid-interactant interlocution: (a) the duration of utterances spoken by confederate interviewees, (b) the difficulty of questions posed by participants, and (c) the sophistication of responses given by confederate interviewees in reply to participants' questions. Our interest resided in whether or not confederates would produce equivalent amounts of content across conditions, whether participants would ask more difficult questions of an adult-bodied interviewee independent of which confederate was actually generating responses to their questions, and also if self-stereotyping would occur, whereby confederates would alter the sophistication of their responses on account of the body they were interviewed through. Furthermore, we assessed participants' written evaluations of the person they interviewed to gauge whether participants' explicit impressions were mediated by the age group of the body they encountered.

Method. Subjects and confederates. Seventy-two adults partook in the study (43 female; mean age = 23.33; $SD = 2.80$). A 12-year-old male actor and a 37-year-old adult male social psychology professor served as confederates. Participants were randomly assigned to interview-panels nested within experimental conditions. As with Study 1, the confederates rehearsed Study 2's procedure in numerous mock trials so as to achieve consistency with the cyranoid technique.

Conditions. Each of four experimental conditions consisted of 18 participants divided among four interview panels, and panels varied in size between three and five participants. Conditions followed an identical protocol wherein interview-panels interrogated a confederate for 20-minutes. In the {[Adult]Adult} condition, interview-panels interacted with the adult confederate speaking autonomously: {[Adult]Adult} \longleftrightarrow Panel₁₋₄. The

{[Adult]Child} condition featured the adult confederate sourcing for the child confederate shadower: {[Adult]Child} \longleftrightarrow Panel₅₋₈. The {[Child]Child} condition consisted of interview-panels interacting with the autonomously-speaking child confederate: {[Child]Child} \longleftrightarrow Panel₉₋₁₂. Finally, the {[Child]Adult} condition featured the child confederate sourcing for the adult confederate shadower: {[Child]Adult} \longleftrightarrow Panel₁₃₋₁₆. Experimental trials alternated so as to counterbalance sequencing effects.

Instructions and materials. Participants received instruction forms individually and were given verbal instructions by the researcher as an interview-panel. Participants were asked to interview an individual in order to gain a sense of “what they’re like and what they know.” They were told to focus their questions on the domains of (a) science, (b) literature, and (c) current and historical political events, these being quite similar to those Milgram (1992) had instructed his participants to follow. The researcher emphasized that they were free to interpret these domains as broadly as they wished. As with Study 1, the emphasis on allowing participants to generate their own questions was intended to undermine the possibility of participants assuming their interviewee’s responses were rehearsed. Participants were asked not to speak to each other during the interview nor respond to any question or comment posed by another panel member (so as to reduce their influence over one another and to keep dialogue directed toward the confederate interviewee).

A post-interview questionnaire completed by participants contained the same three agree/disagree items from Study 1 designed to gauge whether participants succumbed to the cyranic illusion. As with Study 1, participants were also asked to provide a brief written evaluation of their interviewee before being interviewed by the researcher.

Apparatus. The apparatus used was similar to that of Study 1, except that the interaction room contained five chairs positioned opposite the confederate such that each interviewer sat facing their interviewee at a distance of roughly 6 feet. The contraption of devices that allowed the source confederate to deliver speech to the shadowing confederate in cyranic conditions was identical to that utilized in Study 1.

Procedure. Following informed consent and instruction, the confederate (either the adult or the child) was brought into the interaction room and seated opposite the interview-panel. In all conditions the confederate went by the name “Stanley.” Participants were given no

background information on Stanley nor were they given any indication that this person was a confederate. The researcher then left the room and interviews commenced.

The researcher returned after 20-minutes to halt the interview. The confederate left the interaction room and participants were led to an evaluation room, seated at individual desks, and handed the post-interaction questionnaire. Following completion of questionnaires and written evaluations, participants returned to the interaction room where they were interviewed and debriefed by the researcher as a group akin to Study 1.

Interlocution measures. Dialogue from each interview was transcribed and a survey containing all main interviewer question threads extracted from each interview-panel was compiled using survey software and presented to six coders (four female, mean age = 24.00) blind to the research objectives. Only questions which introduced a new topic or concept were assessed, while follow-up questions and comments made by the interviewers that did not significantly change the topic or introduce a new concept were excluded from the analysis. Coders independently rated each question (condition-blind, randomized) in terms of “how difficult to answer the average person would find the question” using a five-point rating system ranging from 1 (*not at all difficult*) to 5 (*very difficult*). The composite variable *Question Difficulty* was derived by averaging the difficulty scores provided by coders for each question and was used to assess whether or not the questions posed by participants varied in terms of difficulty across experimental conditions.

Confederate interviewees’ full responses to the question threads posed by participants (including responses to follow-up questions) were extracted from each experimental trial and presented to five coders (three female, mean age = 24.00) blind to the research objectives. Coders independently rated each response (condition-blind, randomized) in terms of its sophistication on a scale ranging from 1 (*not at all sophisticated*) to 5 (*very sophisticated*). The composite variable *Response Sophistication* was then computed by averaging coders’ sophistication ratings for each response and was used to assess whether or not the sophistication of confederate interviewees’ responses differed across experimental conditions. We also computed the total number of words spoken by the confederate interviewee during each full response in order to generate the variable *Response Length*. This variable was calculated in order to assess the degree to which *Response Sophistication* varied in relation to the length of responses provided by confederate interviewees, as well as to gain

a sense of whether or not perceptions of sophistication varied more as a function of the quantity of words spoken vs. the quality of the words (i.e., content, reasoning, lexicon, phrase structure, etc.).

Finally, the variable *Utterance Length* was generated by calculating the quantity of words articulated during each unique conversational turn spoken by confederate interviewees. This variable was used to determine the extent to which confederates produced similar amounts of turn content when speaking autonomously compared to when sourcing for a shadowing confederate.

Person perception measures. Following all trials, participants' post-interaction written evaluations were transcribed and anonymized. Two independent reviewers (one female, mean age = 28.50), blind to the research objectives, developed a coding frame comprised of seven dichotomous trait dimensions which emerged from the corpus of written evaluations, these being: (a) *Intelligence* (intelligent vs. unintelligent), (b) *Confidence* (confident vs. unconfident), (c) *Maturity* (mature vs. immature), (d) *Extraversion* (extraverted vs. introverted), (e) *Friendliness* (friendly vs. unfriendly), (f) *Opinionatedness* (opinionated vs. not-opinionated), and (g) *Honesty* (honest vs. dishonest). Six coders (four female, mean age = 23.00) then independently rated each participant's written evaluation (condition-blind, randomized), assigning to each a score for each trait dimensions: +1 for positive attributions, -1 for negative attributions, and 0 for attributions that did not appear in the written evaluation (e.g., if a participant remarked that their interviewee was intelligent, confident, mature, introverted, unfriendly, and opinionated, but did not comment on their honesty, a coder would score their evaluation as follows: *Intelligence* = 1, *Confidence* = 1, *Maturity* = 1, *Extraversion* = -1, *Friendliness* = -1, *Opinionatedness* = 1, *Honesty* = 0). For each written evaluation, seven person perception measure variables (*Intelligent*, *Confident*, *Mature*, *Extraverted*, *Friendly*, *Opinionated*, and *Honest*) were generated by averaging coders' trait dimension scores.

Results. Detecting the cyranic illusion. We again used exact logistic regression to analyze participants' post-interaction questionnaire agree/disagree responses. The dummy variables *Cyranoid* and *Adult* were used to signify both the type of interaction experienced by the participants (cyrano vs. non-cyrano) and the confederate present in the room with participants (adult vs. child), respectively. Neither *Cyranoid*, OR = 0.71, SE = 0.59, 95% CI

[0.08, 5.02], *Adult*, OR = 3.29, $SE = 3.93$, 95% CI [0.23, 188.78], nor the interaction between these factors, OR = 0.47, $SE = 0.65$, 95% CI [0.01, 13.05] significantly affected the odds of a participant agreeing with the statement: “The person I interviewed spoke their own thoughts.” Likewise, neither *Cyranoid*, OR = 1.58, $SE = 1.53$, 95% CI [0.16, 21.40], *Adult*, OR = 1.00, $SE = 1.05$, 95% CI [0.06, 15.39], nor the interaction between these factors, OR = 1.00, $SE = 1.31$, 95% CI [0.03, 29.60] significantly affected the odds of agreeing with the statement: “The person I interviewed gave scripted responses.” A model was not calculated for the item stating: “The person I interviewed spoke by receiving radio messages and repeating them,” as exactly one participant ($n = 1$) in each experimental condition agreed with this item while the remaining participants ($n = 17$) in each condition disagreed.

An assessment of the post-interaction written evaluations, interview statements and debriefing remarks revealed strong evidence for the success of the illusion. No participant commented that they suspected that their interviewee had been merely ventriloquizing for another individual, and although a small number of participants suggested that their interviewee might have prepared answers, none felt strongly that the interviewee was behaving in an inauthentic manner. Furthermore, assessments of the video footage and transcripts from each trial showed that not once during any of the 16 interview sessions did participants openly question the interviewee’s autonomy.

Interlocution. Following Shrout and Fleiss’ (1979) guidelines on computing intraclass correlation, high agreement was found among coders who scored the difficulty of questions posed by participants: ICC(2,6), absolute = 0.85, 95% CI [0.80, 0.89]. Similarly, high agreement was found among coders who rated the sophistication of confederate interviewees’ responses: ICC(2,5), absolute = 0.90, 95% CI [0.87, 0.92].

We used procedures demonstrated by Field (2009) and Judd (2000) to build multilevel mixed effects linear regression models in order to assess the contrasts between the experimental conditions with respect to *Utterance Length*, *Question Difficulty*, and *Response Sophistication*. As observations of these interlocution measures were drawn from interview-panels nested within experimental conditions (and therefore non-independent in nature), the random effects of each interview-panel were considered in our models while the dummy variables *[Adult]* and *{Adult}* were used to indicate fixed factor levels pertinent to each experimental condition. *[Adult]* took the value of 1 in conditions where the adult confederate

generated responses to interview-panel questions and 0 in conditions where the child confederate generated responses. $\{\text{Adult}\}$ took the value of 1 when the adult confederate was physically present in the interaction room and 0 when the child confederate was physically present in the interaction room. Thus, the factor levels for each experimental condition were as follows: $\{\text{[Adult]Adult}\}: \text{[Adult]} = 1, \{\text{Adult}\} = 1$; $\{\text{[Adult]Child}\}: \text{[Adult]} = 1, \{\text{Adult}\} = 0$; $\{\text{[Child]Child}\}: \text{[Adult]} = 0, \{\text{Adult}\} = 0$; $\{\text{[Child]Adult}\}: \text{[Adult]} = 0, \{\text{Adult}\} = 1$.

Utterance length. Our final model in which *Utterance Length* was designated as the response variable revealed a significant interaction between $\{\text{Adult}\}$ and [Adult] , $F(7.92) = 9.52, p < .05$. There were no significant main effects of either [Adult] , $F(7.68) = 2.01, p = .20$, or $\{\text{Adult}\}$, $F(7.72) = 0.03, p = .87$. See Table 2.1 for final model estimated effect sizes, standard errors, and confidence intervals, and Table 2.4 for pooled means and standard deviations.

Two separate multilevel models were used to further explore the significant interaction between $\{\text{Adult}\}$ and [Adult] with respect to *Utterance Length*. We first considered only observations from experimental conditions in which participants engaged with the adult confederate's words ($\text{[Adult]} = 1$) and fit a modified version of our final model that excluded [Adult] and the interaction between [Adult] and $\{\text{Adult}\}$ as fixed factors. This model showed that the adult confederate's utterances were significantly longer when spoken through their own body ($\{\text{Adult}\} = 1$) than when shadowed by the child confederate, $b = 66.67, SE = 20.70, t(3.99) = 3.22, p < .05$. We then considered only participants who engaged with the child confederate's words ($\text{[Adult]} = 0$) and found that the child confederate's utterances were not significantly different when being shadowed by the adult confederate ($\{\text{Adult}\} = 1$) than when spoken through their own body, $b = 2.12, SE = 2.54, t(7.72) = 0.83, p = .43$.

Question difficulty. Our final model predicting *Question Difficulty* showed significant main effects of both $\{\text{Adult}\}$, $F(11.97) = 9.18, p < .05$, and [Adult] , $F(15.46) = 6.15, p < .05$. The interaction between these two fixed factors was not significant, $F(17.92) = 1.61, p = .22$. We included in our final model the fixed factor predictor *Previous Response Sophistication* as a control variable in order to examine the relationship between the difficulty of questions posed and the sophistication (*Response Sophistication*) confederate interviewees demonstrated in the preceding question thread. *Previous Response Sophistication* was found to have no significant effect on *Question Difficulty*, $F(362.97) = 1.35, p = .25$. Estimated effect sizes, standard errors, and confidence intervals for both our final and partial models are shown in

Table 2.1

Multilevel Model Fixed Factor Effect Size Estimates for Utterance Length – Study 2^a

Predictors	b (SE)	95% CI
Constant	13.36 (2.16)**	8.14, 18.58
Categorical Fixed Factors ^b :		
{Adult}	2.54 (14.52)	-31.16, 36.25
[Adult]	4.56 (3.22)	-2.91, 12.02
Interaction: {Adult} x [Adult]	63.76 (20.67)*	16.01, 111.51

Note. Dependent measure is *Utterance Length*, defined as number of words per confederate interviewee utterance. Model includes random intercepts and random slope coefficients as the relationship between *{Adult}* and *Utterance Length* showed significant variance in intercepts across interview-panels, $\text{var}(v_{0j}) = 8.37$, $\chi^2(1) = 304.54$, $p < 0.001$, while slopes also varied significantly across interview-panels, $\text{var}(v_{1j}) = 800.74$, $\chi^2(1) = 23.70$, $p < 0.001$.

^a {[Adult]Adult}: $n = 139$, 4 panels; {[Child]Adult}: $n = 262$, 4 panels; {[Child]Child}: $n = 402$, 4 panels; {[Adult]Child}: $n = 290$, 4 panels.

^b Factor levels for {[Child]Child}: {Adult} = 0, [Adult] = 0; {[Child]Adult}: {Adult} = 1, [Adult] = 0; {[Adult]Child}: {Adult} = 0, [Adult] = 1; {[Adult]Adult}: {Adult} = 1, [Adult] = 1.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Table 2.2, while Table 2.4 shows pooled means and standard deviations.

Response sophistication. Our final model with the response variable *Response Sophistication* revealed a significant main effect of [Adult], $F(11.22) = 21.18$, $p < .01$, a non-significant main effect of {Adult}, $F(10.98) = 0.42$, $p = .53$, and a significant interaction between these factors, $F(12.94) = 7.00$, $p < .05$. Estimated effect sizes, standard errors, and confidence intervals can be found in Table 2.3, while Table 2.4 shows pooled means and standard deviations.

Separate multilevel models were used to explore the interaction between {Adult} and [Adult]. We first considered only observations from experimental conditions in which participants engaged with the adult confederate's words ([Adult] = 1) and fit a modified version of our final model that excluded [Adult] and the interaction between [Adult] and {Adult} as fixed factors. This model revealed that the effect of {Adult} on response sophistication scores was significant, $b = 1.00$, $SE = 0.30$, $t(6.81) = 3.36$, $p < .05$. We then considered only participants who engaged with the child confederate's words ([Adult] = 0) and found no significant effect of {Adult}, $b = 0.14$, $SE = 0.15$, $t(7.02) = 0.89$, $p = .40$. As a post-hoc analysis, we decided to explore the possibility that relative differences in *Response Length* (the total words spoken by a confederate during their complete response to a question thread) may have played a central

Table 2.2

Multilevel Model Fixed Factor Effect Size Estimates for Question Difficulty – Study 2^a

Predictors	Partial Model		Final Model ^b	
	b (SE)	95% CI	b (SE)	95% CI
Constant	2.01 (0.15)***	1.69, 2.34	1.88 (0.18)***	1.50, 2.25
Categorical Fixed Factors ^c :				
{Adult}	0.64 (0.21)*	0.17, 1.11	0.63 (0.21)*	0.18, 1.08
[Adult]	0.63 (0.22)*	0.16, 1.10	0.55 (0.22)*	0.08, 1.02
Interaction: {Adult} x [Adult]	-0.34 (0.33)	-1.04, 0.36	-0.41 (0.33)	-1.10, 0.27
Control Variable:				
Previous Response			0.08 (0.07)	-0.06, 0.22
Sophistication ^d				

Note. Dependent measure is *Question Difficulty*, computed by averaging coder difficulty ratings for each question posed by interview-panel members (ranges from 1 to 5: 1 = not at all difficult; 5 = very difficult). Final Model includes random intercepts as the relationship between {Adult} and *Question Difficulty* showed significant variance in intercepts across interview-panels, $\text{var}(v_{0j}) = 0.06, \chi^2(1) = 11.41, p < 0.001$.

^a {[Adult]Adult}: $n = 39$, 4 panels; {[Child]Adult}: $n = 91$, 4 panels; {[Child]Child}: $n = 158$, 4 panels; {[Adult]Child}: $n = 78$, 4 panels.

^b Final Model (-2 log-likelihood = 958.67) showed slight improvement in fit vs. Partial Model (-2 log-likelihood = 959.99).

^c Factor levels for {[Child]Child}: {Adult} = 0, [Adult] = 0; {[Child]Adult}: {Adult} = 1, [Adult] = 0; {[Adult]Child}: {Adult} = 0, [Adult] = 1; {[Adult]Adult}: {Adult} = 1, [Adult] = 1.

^d Defined as *Response Sophistication* score for confederate interviewee's response to preceding question.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Table 2.3

Multilevel Model Fixed Factor Effect Size Estimates for Response Sophistication – Study 2^a

Predictors	b (SE)	95% CI
Constant	1.75 (0.15)***	1.41, 2.09
Categorical Fixed Factors ^b :		
{Adult}	0.14 (0.22)	-0.34, 0.62
[Adult]	1.01 (0.22)**	0.52, 1.50
Interaction: {Adult} x [Adult]	0.85 (0.32)*	0.16, 1.55

Note. Dependent measure is *Response Sophistication*, computed by averaging coder sophistication ratings for each statement made by confederate interviewee in response to interview-panel members' questions (ranges from 1 to 5: 1 = not at all sophisticated; 5 = very sophisticated). Final Model includes random intercepts as the relationship between {Adult} and *Response Sophistication* showed significant variance in intercepts across interview-panels, $\text{var}(v_{0j}) = 8.00 \times 10^{-2}, \chi^2(1) = 21.36, p < .001$.

^a {[Adult]Adult}: $n = 43$, 4 panels; {[Child]Adult}: $n = 95$, 4 panels; {[Child]Child}: $n = 162$, 4 panels; {[Adult]Child}: $n = 82$, 4 panels.

^b Factor levels for {[Child]Child}: {Adult} = 0, [Adult] = 0; {[Child]Adult}: {Adult} = 1, [Adult] = 0; {[Adult]Child}: {Adult} = 0, [Adult] = 1; {[Adult]Adult}: {Adult} = 1, [Adult] = 1.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Table 2.4

Comparison of Interlocution Measures – Study 2

Condition	Utterance Length ^a		Question Difficulty ^b		Response Sophistication ^c	
	n	M(SD) ^d	n	M(SD) ^d	n	M(SD) ^d
{[Adult]Adult}	139	72.09 (85.41)	39	2.91 (1.07)	43	3.68 (1.09)
{[Adult]Child}	290	17.31 (18.26)	78	2.63 (0.92)	82	2.74 (0.79)
{[Child]Adult}	262	15.26 (15.48)	91	2.64 (0.98)	95	1.84 (0.62)
{[Child]Child}	402	13.03 (14.56)	158	2.04 (0.83)	162	1.70 (0.58)

Note. Table displays pooled means for interlocution measures drawn from each experimental condition. For multilevel statistical comparisons of fixed factor effect size estimates and significance testing, see Tables 1, 2, and 3.

^a *Utterance Length* defined as number of words per confederate interviewee utterance.

^b *Question Difficulty* computed by averaging coder difficulty ratings for each question posed by interview-panel members; ranges from 1 to 5 (1 = not at all difficult; 5 = very difficult).

^c *Response Sophistication* computed by averaging coder response sophistication ratings for each full statement made by confederate interviewee in response to interview-panel members' questions; ranges from 1 to 5 (1 = not at all sophisticated; 5 = very sophisticated).

^d Means and associated standard deviations derived by pooling each observation by experimental condition (ignores interview-panel nesting).

role in the disparity in response sophistication scores between {[Adult]Adult} and {[Adult]Child}, as the adult confederate produced significantly lengthier responses when speaking through their own body than when speaking through the body of the child confederate, $b = 86.03$, $SE = 12.00$, $t(123) = 7.19$, $p < .001$. Indeed, considering only cases in which the adult confederate generated responses ([Adult] = 1), *Response Sophistication* was significantly predicted by *Response Length*, $b = 0.10$, $SE = 1.00 \times 10^{-14}$, $t(123) = 12.56$, $p < .001$.

Person perception. Inter-rater reliability analyses showed high agreement for each of the seven coded trait dimensions derived from participants' post-interaction written evaluations (*Intelligence*: ICC(2,6), absolute = 0.92, 95% CI [0.88, 0.94]; *Confidence*: ICC(2,6), absolute = 0.86, 95% CI [0.80, 0.91]; *Maturity*: ICC(2,6), absolute = 0.83, 95% CI [0.76, 0.88]; *Extraversion*: ICC(2,6), absolute = 0.80, 95% CI [0.72, 0.87]; *Friendliness*: ICC(2,6), absolute = 0.82, 95% CI [0.74, 0.88]; *Opinionatedness*: ICC(2,6), absolute = 0.80, 95% CI [0.72, 0.87]; *Honesty*: ICC(2,6), absolute = 0.84, 95% CI [0.77, 0.89]).

{[Adult]Adult} vs. {[Adult]Child}. To gain a sense of how participants' perceptions of the

autonomously communicating adult confederate compared to perceptions of the adult confederate speaking through the body of a child, we computed pooled means for each person perception measure (*Intelligent, Confident, Mature, Extraverted, Friendly, Opinionated, and Honest*) and conducted independent samples means tests contrasting {[Adult]Adult} with {[Adult]Child}. For *Mature*, scores from the {[Adult]Child} condition were significantly greater than those from the {[Adult]Adult} condition, $t(34) = 2.25, p < .05$, signifying that participants in the {[Adult]Child} condition commented on their interviewee's maturity in their post-interaction written evaluations significantly more so than participants in the {[Adult]Adult} condition. No significant differences were found with regard to *Intelligent*, $t(34) = 1.72, p = .09$, *Confident*, $t(34) = -0.62, p = .54$, *Extraverted*, $t(34) = 0.19, p = .85$, *Friendly*, $t(34) = 0.16, p = .88$, *Opinionated*, $t(34) = -1.51, p = .14$, and *Honest*, $t(34) = 0.68, p = .50$. Pooled means and standard deviations for each person perception measure are shown in Table 2.5.

The following are excerpts from participants' post-interaction written evaluations of {[Adult]Adult}. In general, participants found the autonomously speaking adult to be intelligent and engaging:

Stanley was obviously very intelligent, at many points I couldn't even follow his well thought-out arguments which tended to be quite philosophical. He obviously had at least a basic knowledge of all 3 subject areas, if not a very deep understanding of the issues surrounding them. ([Adult]Adult), Panel 1, participant 3; person perception measure scores: *Intelligent* = 1.00, *Confident* = 0.50, *Mature* = 0.50, *Extraverted* = 0.17, *Friendly* = 0.00, *Opinionated* = 0.17, *Honest* = 0.00).

Stanley is a very knowledgeable person. He is very aware of current political and historic issues. In addition, he is capable of expressing deep expertise in a friendly manner. I would definitely consult him for advice in economics and history. ([Adult]Adult), Panel 2, participant 8; person perception measure scores: *Intelligent* = 1.00, *Confident* = 0.67, *Mature* =

Table 2.5

Person perception score comparison – Study 2: {[Adult]Adult} vs. {[Adult]Child}

Person Perception Dimension	{[Adult]Adult}	{[Adult]Child}	Δ
	M (SD) ^{a, b}	M (SD) ^{a, b}	
Intelligent	0.77 (0.30)	0.94 (0.27)	0.17
Confident	0.58 (0.24)	0.52 (0.38)	-0.06
Mature	0.35 (0.26)	0.59 (0.37)	0.24*
Extraverted	0.30 (0.28)	0.31 (0.30)	0.01
Friendly	0.25 (0.37)	0.27 (0.33)	0.02
Opinionated	0.43 (0.37)	0.21 (0.47)	-0.22
Honest	0.13 (0.32)	0.20 (0.34)	0.07

Note. Table compares pooled mean person perception scores attributed to confederate interviewee in participants' post-interaction written evaluations between {[Adult]Adult} and {[Adult]Child}. Differences in means evaluated using independent samples t-tests.

^a Means and associated standard deviations derived by pooling each person perception score by experimental condition (ignores interview-panel nesting).

^b The possible range for mean scores is between -1 and +1, with positive scores reflecting more frequent positive evaluation of confederate interviewee in participants' post-interaction written evaluations (e.g., describing the interviewee as "intelligent"), and negative scores reflecting more frequent negative evaluation of confederate (e.g., describing the interviewee as "unintelligent").

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

0.33, *Extraverted* = 0.17, *Friendly* = 1.00, *Opinionated* = 0.50, *Honest* = 0.17).

I think he is an honest person and talkative. Sometimes even if he doesn't know for sure about the answer he can still find something to talk about to get the conversation going. ({[Adult]Adult}, Panel 4, participant 18; person perception measure scores: *Intelligent* = 0.33, *Confident* = 0.67, *Mature* = 0.50, *Extraverted* = 1.00, *Friendly* = 0.50, *Opinionated* = 0.17, *Honest* = 1.00).

The following are excerpts from participants' post-interaction written evaluations of {[Adult]Child}. It is interesting to note that while participants were largely impressed by the intelligence of the person with whom they interacted, some qualify their evaluations by suggesting that their interviewee may not have fully understood what they were

communicating (see second and third excerpts below):

Very intelligent. Eloquent and charming. He is obviously very bright and has a very high level of knowledge for his age. He was very polite and well-mannered. He handled a slightly odd social situation very maturely. ([Adult]Child}, Panel 8, participant 32; person perception measure scores: *Intelligent* = 1.00, *Confident* = 0.67, *Mature* = 1.00, *Extraverted* = 0.33, *Friendly* = 0.67, *Opinionated* = 0.50, *Honest* = 0.50).

Clearly very bright. His ability to connect patterns and think in a more expanded way than some the questions implied was more impressive than his name-dropping of the classics (although I don't doubt that he read and understood them). He was articulate, with a lot of intellectual potential, especially regarding linkages between language and knowledge. I thought he'd enjoy discourse analysis at some point. ([Adult]Child}, Panel 5, participant 19; person perception measure scores: *Intelligent* = 1.00, *Confident* = 0.83, *Mature* = 0.83, *Extraverted* = 0.33, *Friendly* = 0.00, *Opinionated* = 0.66, *Honest* = 0.00).

He seemed rather intelligent for his age, with a lot of knowledge on various matters. He seemed to be capable of using big words, however I wasn't certain if he actually knew what they meant in context. ([Adult]Child}, Panel 6, participant 25; person perception measure scores: *Intelligent* = 1.00, *Confident* = 0.33, *Mature* = 0.33, *Extraverted* = 0.00, *Friendly* = 0.00, *Opinionated* = 0.33, *Honest* = -0.33).

{[Child]Child} vs. {[Child]Adult}. We again calculated pooled means for each person perception measure, this time comparing scores from the {[Child]Child} condition with those from the {[Child]Adult} condition. Scores from {[Child]Child} were significantly greater than those from {[Child]Adult} with regard to *Intelligent*, $t(34) = 7.37$, $p < .001$, *Confident*,

$t(34) = 5.22, p < .001$, *Mature*, $t(34) = 3.60, p < .01$, *Extraverted*, $t(34) = 3.17, p < .01$, *Friendly*, $t(34) = 3.08, p < .01$, and *Opinionated*, $t(34) = 4.63, p < .001$. Only with respect to *Honest* did scores not significantly differ, $t(34) = 1.82, p = .08$. Pooled means and standard deviations for each person perception measure are displayed in Table 2.6.

The following are excerpts from participants' post-interaction written evaluations of {[Child]Child}. The interviewee in this condition received generally quite positive evaluations, while many participants explicitly referenced the interviewee's youth and age:

Very willing to venture a guess about questions asked. Honest about his uncertainty on some topics. Shows maturity. Well informed on local current issues and knows a wide range of information about basic science, which he says is his favourite topic. He mentioned one author he enjoyed but doesn't seem interested in reading literature. Smart, at the same level I would expect of his age." ({[Child]Child}, Panel 10, participant 43; person perception measure scores: *Intelligent* = 1.00, *Confident* = 0.50, *Mature* = 1.00, *Extraverted* = 0.33, *Friendly* = 0.17, *Opinionated* = 0.67, *Honest* = 0.83).

I thought he was a friendly lad, and as expected he didn't know lots about literature, science, etc... Open to ideas, able to consider new points of view. Seemed intelligent. ({[Child]Child}, Panel 11, participant 48; person perception measure scores: *Intelligent* = 1.00, *Confident* = 0.17, *Mature* = 0.33, *Extraverted* = 0.33, *Friendly* = 0.67, *Opinionated* = 0.17, *Honest* = 0.33).

Cute. Funny. Knowledgeable for a kid, has a good interest for science as well as current events. ({[Child]Child}, Panel 12, participant 51; person perception measure scores: *Intelligent* = 1.00, *Confident* = 0.50, *Mature* = -0.17, *Extraverted* = 0.83, *Friendly* = 0.83, *Opinionated* = 0.17, *Honest* = 0.00).

Table 2.6

Person perception score comparison – Study 2: {[Child]Child} vs. {[Child]Adult}

Person Perception Dimension	{[Child]Child}	{[Child]Adult}	Δ
	M (SD) ^{a, b}	M (SD) ^{a, b}	
Intelligent	0.68 (0.39)	-0.37 (0.46)	-1.05***
Confident	0.31 (0.39)	-0.41 (0.43)	-0.72***
Mature	0.21 (0.48)	-0.26 (0.28)	-0.47**
Extraverted	0.19 (0.40)	-0.22 (0.39)	-0.41**
Friendly	0.30 (0.33)	-0.05 (0.34)	-0.35**
Opinionated	0.26 (0.43)	-0.38 (0.40)	-0.64***
Honest	0.30 (0.34)	0.08 (0.37)	-0.22

Note. Table compares pooled mean person perception scores attributed to confederate interviewee in participants' post-interaction written evaluations between {[Child]Child} and {[Child]Adult}. Differences in means evaluated using independent samples t-tests.

^a Means and associated standard deviations derived by pooling each person perception score by experimental condition (ignores interview-panel nesting).

^b The possible range for mean scores is between -1 and +1, with positive scores reflecting more frequent positive evaluation of confederate interviewee in participants' post-interaction written evaluations (e.g., describing the interviewee as "intelligent"), and negative scores reflecting more frequent negative evaluation of confederate (e.g., describing the interviewee as "unintelligent").

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

The following are excerpts from participants' post-interaction written evaluations of {[Child]Adult}. We can see that despite the fact that the same child confederate generated responses to interviewee's questions in both conditions, participants evaluated {[Child]Child} much more favorably than {[Child]Adult} (with whom participants were largely unimpressed):

I feel Stanley was quite an aloof in the matters of science, literature and current/hist. political events. He lacked the ability to expand on any points he mentioned. Although, he did seem to be quite passionate about graphic novels. Overall, it feels as if Stanley was in a world of his own. His actions were quite hesitant. ({[Child]Adult}, Panel 13, participant 58; person perception measure scores: *Intelligent* = -0.67, *Confident* = -0.50, *Mature* = -0.33, *Extraverted* = -0.83, *Friendly* = -0.67, *Opinionated* = -0.67, *Honest* = 0.00).

He did not have very much to say regarding the topics given. The way he expressed his opinions was a bit unorganized. I wasn't sure if he was informed about the topics and unable to express himself or if he didn't understand the topics. He gave off a rather adult impression at first but as the conversation went on seemed a bit uninterested in most of the topics. ([Child]Adult), Panel 16, participant 71; person perception measure scores: *Intelligent* = -0.83, *Confident* = -0.67, *Mature* = -0.67, *Extraverted* = -0.50, *Friendly* = -0.17, *Opinionated* = -0.67, *Honest* = 0.00).

He has knowledge of a high school graduate. He does not have the capability of thinking critically. ([Child]Adult), Panel 13, participant 55; person perception measure scores: *Intelligent* = -0.67, *Confident* = -0.17, *Mature* = -0.67, *Extraverted* = 0.00, *Friendly* = 0.00, *Opinionated* = -0.50, *Honest* = 0.00).

General discussion

Findings. In both studies, participants' written and debrief statements provided perhaps the strongest evidence in favor of the cyranic illusion, though analyses of participants' responses to agree/disagree questionnaire items also proved convincing. Even in cases involving significant age incongruity between source and shadower (Study 2), participants failed to notice when their interlocutor was not self-authoring the words he spoke, suggesting that the cyranic illusion is a robust phenomenon not limited to instances of high source-shadower congruence. It seems that when encountering an interlocutor face-to-face, people rarely question whether the "mind" and the "body" of a person are indeed unified—and for good reason, as social interaction would be undermined if we began to doubt whether each person we encountered was indeed the true author of the words they expressed. This observation regarding everyday social life stands in contrast to socialization that occurs in artificial environments (e.g., Second Life and other virtual community games) wherein users can construct outer personae which starkly contrast with their real-world identities (see Bessiere, Seay, & Kiesler, 2007; Vasalou & Joinson, 2009), and wherein unity between the user and their avatar is more readily questioned (see Donath, 1999).

Our analysis of *Utterance Length* in Study 2 showed that {[Adult]Adult} generated far greater utterance lengths than {[Adult]Child}, {[Child]Adult} and {[Child]Child}, which with respect to *Utterance Length* were statistically equivalent. There are two immediate explanations for why such a discrepancy may have occurred. First, it is entirely possible that the adult confederate's behavior confirmed stereotyped assumptions about the linguistic limitations (e.g., reduced ability to expand upon complex concepts) expected of the child confederate relative to their own, thus resulting in shortened prose when sourcing. This would suggest that the child confederate may have been capable of shadowing much longer utterances but was simply not afforded the chance to do so. The fact that a corollary pattern did not emerge when the child, in turn, sourced words for the adult confederate may suggest that whereas adults perhaps have the ability to alter their verbal behavior so as to speak utterance lengths akin to those typically used by children, children may not as readily be able to generate utterances as lengthy as those produced by adults (particularly university professors). However, this discrepancy is most likely an artifact born of the unique relationship between the confederates used in our study, as the adult confederate found it much easier to source abbreviated passages than paragraph-length prose and suggested that at times it was difficult for the child to accurately shadow long and complex speech. This implies that the cyranoid method might be constrained by functional factors, namely that certain source-shadower pairings may require that a source openly alter certain aspects of their verbal behavior in order for the cyranoid to function in a manner that preserves the illusion of autonomy. It is a limitation of our study that we did not more closely account for how and why differences in speech shadowing ability may have altered utterance lengths, though we recognize that source-shadower functional impediments are certainly worth investigating in their own right in future research.

In considering the final model for our *Question Difficulty* interlocution measure, the results of Study 2 provide mixed evidence for the notion that people's verbal behavior will confirm age-based stereotypes during social interaction. Our results indicate that encountering adult-generated responses (*[Adult] = 1*) overrode participants' inclination to ask "child-level" questions when faced with a child-bodied interviewee in that participants' questions in {[Adult]Child} were significantly more difficult than those in {[Child]Child}. On the other hand, the fact the difficulty of questions in {[Child]Adult} and {[Adult]Adult} were statistically comparable suggests that here participants' verbal behavior *did* confirm stereotyped assumptions about what types of questions one should ask another person on the

basis of their physical age and independent of their actual capacity answer such questions. Considering the literature on situational ambiguity and heuristic processing (e.g., Bohner, Chaiken, & Hunyadi, 1994; Chaiken & Maheswaran, 1994) might help partially resolve this discrepancy. Participants in {[Adult]Child} may have adjusted the difficulty of their questions upward while those who encountered {[Child]Adult} did not seem to make a related adjustment downward due to relative differences in ambiguity between the two contexts. Participants in the {[Child]Adult} condition were presented with a situation in which their interviewee produced far less content than what might have been expected, as utterance lengths in the {[Adult]Adult} condition were significantly greater than those in {[Child]Adult}. The comparative lack of content produced by {[Child]Adult}, therefore, may have generated ambiguity by-way-of a violated expectation, leading participants to anchor on age-based heuristic cues and thus continue to pose questions comparable in difficulty to the {[Adult]Adult} condition. On the other hand, utterance lengths between {[Adult]Child} and {[Child]Child} were statistically similar, therefore the amount of content produced by {[Adult]Child} may not have comprised a violation of expectations, and as such the difficulty level of participants' questions in the {[Adult]Child} more directly tracked the adult-generated words spoken by their interviewee.

An alternative interpretation of the discrepancies in question difficulty across Study 2's conditions involves considering that participants in the {[Child]Adult} condition may have refrained from lessening the difficulty of their questions on account of the fact that doing so might have been perceived as a form of *talking down* to their interviewee. Indeed, narrative researchers have pointed out that it is often quite difficult to speak appropriately with those who display a level of intelligence significantly diminished from what is considered normal (Biklen & Moseley, 1988; Booth & Booth, 1996). It is perhaps more socially acceptable to flatter a child stranger's intelligence by asking them adult-level questions than it is to imply a middle-aged stranger's lack of intelligence by asking them child-level questions, as doing so would contradict the status usually ascribed to members of their age-group (see Barker, Giles, & Harwood, 2004; Harwood, Giles, Clement, Pierson, & Fox, 1994). Moreover, talking down to others is in general a form of condescension, which tends to negatively correlate with prosocial forms of behavior (Nave, Sherman, Funder, Hampson, & Goldberg, 2010).

The analysis of *Response Sophistication* (Study 2) suggests that when sourcing for a shadower, the sophistication of one's verbal communication will not necessarily confirm

stereotypes related to one's perceived physical identity. Our results demonstrate that the sophistication of the child confederate's responses did not differ on the basis of the body-type he communicated through. On the other hand, the adult confederate produced significantly less sophisticated responses when he spoke through the child confederate's body relative to when speaking through his own. This disparity, however, appears to be accounted for by the fact that the adult confederate produced far shorter responses (*Response Length*) when being shadowed by the child confederate compared to when speaking autonomously, with shorter response lengths strongly predicting lower ratings of response sophistication. However, as noted above with respect to *Utterance Length*, differences in *Response Length* between {[Adult]Adult} and {[Adult]Child} were more likely a result of the functional speech shadowing limitations of the child confederate than any sort of self-stereotyping on the part of the adult confederate. Functional limitations such as this aside, the results from Study 2 show that it is possible for sources to maintain a consistent disposition when communicating through a shadower they know to be physically incongruent from themselves. This is perhaps a result of the training procedures we employed, as confederates had time to rehearse the procedures and settle into "being themselves" while sourcing. Had we instead assigned participants without knowledge of the research objectives to function as sources across experimental conditions, we may have observed source behavior confirm stereotypes held by both themselves and interactants. This is indeed an avenue that warrants future investigation considering the literature on self-stereotyping, social cues, and behavioral confirmation (e.g., Chen & Bargh, 1997; Wheeler & Petty, 2001), particularly that which regards behavioral alignment with avatar-identity in virtual environments (e.g., Fox, Bailenson, & Tricase, 2013; Peña, Hancock, & Merola, 2009; Yee et al., 2009).

The cyranoid technique proved a novel route to exploring the relationship between person perception and outer vs. inner identity. Despite the same child confederate generating responses in both conditions, significant divergence occurred between {[Child]Child} and {[Child]Adult} on nearly every trait dimension captured by our coding frame, with {[Child]Adult}, on average, being more negatively perceived on all dimensions. Comparing {[Adult]Adult} with {[Child]Child}, we can see that on their own our confederates were evaluated quite similarly and favorably by participants. However, when we constructed hybrid personae, or "mash-ups" of these two characters in cyranoid conditions, perceptions dramatically altered on account of which confederate constituted mind and which constituted body. Interestingly, we see a contrast between the relatively *high* difficulty of questions

posed to {[Child]Adult} in during panel interviews—a public forum—and the quite *low* opinions expressed of {[Child]Adult} in post-interaction written evaluations—a private setting. This pattern is evidence of the frequent tension between private and public expression of attitudes discussed by Moscovici (1976). These results suggest that the cyranoid method holds promise for investigations into how social perceptions are mediated separately by inner disposition and outer appearance, and how the alignment of public and private expressions of perception shift on account of the mixture of identities one encounters. Though the relationship between person perception and outer vs. inner identity has been studied in virtual environments in recent years (see Neff, Wang, Abbott, & Walker, 2010; Nowak & Rauh, 2005), cyranoids present an opportunity to approach these questions using unscripted human interaction in face-to-face settings.

Future research areas. *Focus: Interactants.* The cyranoid method can be used to address classic questions of person perception, principally those that center on how people separately process verbal and nonverbal cues when forming impressions of and subsequently interacting with other people. In particular, the method can potentially extend literature on discrimination and stereotyping. Experimental research has shown that discrimination can operate at the level of implicit attitudes (McConnell & Leibold, 2001; Nosek & Banaji, 2001) manifesting in behaviors that confirm stereotypes. Via the cyranoid technique, researchers can further explore how these implicit stereotypes reveal themselves in face-to-face, unscripted interactions by separately controlling and manipulating the inner and outer identities of cyranoid stimuli by skin color, age, gender, and so on while observing the behaviors of interactants.

The cyranoid method can also facilitate breaching experiments designed to investigate how people perceive those with whom they have close relationships. Mitchel and colleagues' (2011) artistic demonstration involving spouses encountering partners who have assumed alternative physical identities by-way-of a cyranoid provides an interesting thought experiment: we may think that what we connect with when we encounter an intimate partner, relative, or close friend is some essential, historically-consistent, and ethereal personal quality that exists beyond their physical nature. But in reality, waking up to find your partner had completely changed bodies while their memory and behavioral norms were otherwise intact would, in addition to being quite unsettling, serve to accentuate just how much physical presentation underscores our social relationships. Such scenarios encompass what Garfinkel

(1964) referred to as “breached” social environments, within which the breakdown of familiar norms and the ensuing social psychological phenomena that unfold emphasize how dependent social life is upon mundane expectations.

Focus: Sources. Other potential research questions arise when we turn our attention to the position of the source in cyranic interactions. For instance, it would be particularly worthwhile to study whether biases toward certain person-categories (e.g., race, gender, age, etc.) attenuate or magnify following their sourcing for a shadower whose external identity contrasts markedly from their own. Indeed, the effect of embodying racial- and age-differentiated avatars through the use of immersive virtual environment technology has demonstrated changes in implicit stereotyping and perspective-taking following embodiment (see Groom, Bailenson, & Nass, 2009; Yee & Bailenson, 2006). We suspect that a source’s emotional and empathic responses following cyranoid embodiment might actually be much stronger than those demonstrated in virtual studies given the qualitative realism of a cyranic interaction: cyranoid embodiment involves interacting in-the-flesh with other humans through the body of another human, whereas by comparison immersive virtual environment embodiment of digital avatars is a more mediated and distal experience.

Focus: Shadowers. Finally, the experience of the shadower in cyranic interactions provides another avenue of possibilities. For example, Robb Mitchell and colleagues have explored using shadowers as teaching surrogates in classroom environments (Mitchell, 2010; Raudaskoski & Mitchell, 2013). In these scenarios, remote teachers sourced for student shadowers in classrooms while delivering a lesson, with the role of the shadower rotating amongst the different students. Each student therefore had the opportunity to both present to the class in the form of a cyranoid as well as learn via the variety of teacher-student hybrids possible within the peer group. The authors suggest that the practice of shadowing for teachers in peer activities such as these may help scaffold students’ learning and presentation skills. Along these lines, one could easily imagine the cyranic technique applied to helping those who suffer from social phobias, such as fear of public speaking, learn to overcome their anxieties by allowing sufferers to be guided by experts during social encounters that would otherwise provoke unease. Lastly, the method could be enveloped into current clinical training practices wherein experts can remotely guide the behavior of trainees (see Gordon, 1975).

Conclusion. Though Milgram did not live to see his cyranoid method come to fruition, the current research provides ample basis for the continued exploration of this intriguing methodological paradigm. There are many core domains within social psychology that can be approached with the technique and stand to benefit from the mundane realism that cyranoids bring to the laboratory (not to mention how enjoyable they are for participants to experience). Indeed, the cyranoid method may yet prove to be a long overdue addition to the social psychologist's toolkit.

CHAPTER 3 | Echoborgs: Cyranoids With Computer Program Sources

Preface

Chapter 2 demonstrated how the cyranic illusion can hold under circumstances wherein a source and shadower are quite incongruent from one another. The cyranoid method therefore provides researchers with an interactive and socially dynamic means of studying how a particular human body-type shapes people's perceptions of various dispositional characteristics typically not associated with said body-type, and can do so in the contexts of face-to-face interaction. Moreover, because the identity of the source can remain hidden from interactants over the course of a cyranic encounter, the cyranoid method affords one the ability to not only study situations wherein a source is highly incongruent from their shadower, but also an opportunity to study situations wherein a source isn't even a real person.

The studies contained in Chapter 3 of this thesis pursue exactly these types of scenarios (specifically, situations wherein the source of a cyranoid is a conversational agent computer program that mimics a human interlocutor, creating a special type of cyranoid referred to as an "echoborg"). This maneuver brings the cyranoid method into the world of human-computer interaction research. Though many research domains could have been pursued following the completion of the replication studies described in the previous chapter (e.g., stereotyping research involving cyranoids whose shadowers and sources were differentiated by, say, race or gender), the decision to go down the road of human-computer interaction research was decided upon for five principal reasons:

1. The incongruity between the mind and body of an echoborg is significantly more extreme than that of a mere adult-child cyranoid; thus, exploring echoborg interactions pushes the limits of the cyranic illusion and allows researchers the ability to see whether people continue to perceive mind-body congruence even in the case where the mind is that of a machine.
2. Never before has human-computer interaction research involved situations wherein people encounter the words of a machine face-to-face and in person with an actual human body under conditions wherein they believe they are speaking to an autonomously communicating person. Until now, such scenarios have been the stuff

of science fiction. The echoborg provides a means of investigating how an actual human body, as the perceived originator of linguistic communication, shapes people's perceptions of and experiences with computer-authored words.

3. Echoborgs presented an opportunity to contribute to and complete an incomplete matrix of person stimuli utilized in the field of android science – a discipline that uses machine imitations of humans to study elements of human social psychology and cognition (see MacDorman & Ishiguro, 2006b). Prior to the echoborg, android science research primarily involved experimental comparisons of people's interactions with and perceptions of three general categories of person stimuli: autonomous androids (both mind and body are machine), semi-autonomous androids (human minds controlling machine bodies; e.g., tele-operated androids), and real people (both mind and body are human). The echoborg added the possibility of a person stimulus composed of a machine mind controlling a human body.
4. Echoborgs can be used to investigate how the elements of an interface shape whether or not machine intelligence passes as truly human given certain social psychological conditions (here “passing” refers to the degree to which the identity of a social actor can be taken as authentic; Goffman, 1963; Renfrow, 2004; Khanna & Johnson, 2010).
5. The echoborg concept can contribute to longstanding issues in the realm of philosophy of mind, particularly those tied to Turing's (1950) famous “imitation game” and other thought experiments involving developing third-person/observational criteria for determining the presence or absence of subjective experience in an interlocutor.

The journal article that follows was authored by Corti and Gillespie (2015a). Corti designed and carried out the studies, performed data analysis, outlined the article, and authored the main drafts, contributing roughly 75% of the content. Gillespie provided the technological devices, helped conduct the studies, provided key supervisory assistance, provided funding, provided Corti with editorial suggestions for the article, and contributed roughly 25% of the content. A special section editor at *Frontiers in Psychology*, as well as two anonymous reviewers, experts in the field of android science, provided helpful feedback, suggested edits,

and ultimately approved the article for publication.

Copies of participant instructions and other materials used to conduct the studies in this chapter can be found in Appendix B. Example transcripts from the experimental conditions described in this chapter can be found in Appendix E.

CHAPTER 3 | Echoborgs: Cyranoids With Computer Program Sources

Article Title

A truly human interface: Interacting face-to-face with someone whose words are determined by a computer program

Authors

Kevin Corti (primary)

Alex Gillespie

Abstract

We use speech shadowing to create situations wherein people converse in person with a human whose words are determined by a conversational agent computer program. Speech shadowing involves a person (the shadower) repeating vocal stimuli originating from a separate communication source in real-time. Humans shadowing for conversational agent sources (e.g., chat bots) become hybrid agents (“echoborgs”) capable of face-to-face interlocution. We report three studies that investigated people’s experiences interacting with echoborgs and the extent to which echoborgs pass as autonomous humans. First, participants in a Turing Test spoke with a chat bot via either a text interface or an echoborg. Human shadowing did not improve the chat bot’s chance of passing but did increase interrogators’ ratings of how human-like the chat bot seemed. In our second study, participants had to decide whether their interlocutor produced words generated by a chat bot or simply pretended to be one. Compared to those who engaged a text interface, participants who engaged an echoborg were more likely to perceive their interlocutor as pretending to be a chat bot. In our third study, participants were naïve to the fact that their interlocutor produced words generated by a chat bot. Unlike those who engaged a text interface, the vast majority of participants who engaged an echoborg did not sense a robotic interaction. These findings have implications for android science, the Turing Test paradigm, and human–computer interaction. The human body, as the delivery mechanism of communication, fundamentally alters the social psychological dynamics of interactions with machine intelligence.

Keywords

Android science, cyranoid, dialog system, embodiment, human-computer interaction, speech

shadowing, Turing Test, uncanny valley

Introduction

“Meaning is the face of the Other, and all recourse to words takes place already within the primordial face to face of language” – (Levinas, 1991, p. 206).

In comparison to other forms of interaction, face-to-face communication between humans is characterized by more social emotion, higher demands for comprehensibility, and increased social obligation; the face of the other commands an ethical relation that is absent in people’s interaction with “things” (Levinas, 1991). Face-to-face, close-proximity interaction between tangible bodies is the primordial human inter-face and is the format of exchange most conducive for shared understanding (Linell, 2009). Computer technologies specifically designed to simulate human social functioning (e.g., conversational agents) have to date communicated with people via technical interfaces such as screens, buttons, robotic devices, avatars, interactive voice response systems, and so on. This leaves a need to explore human perception of and interaction with these technologies under conditions that replicate the full complexity of face-to-face human–human communication. The present article introduces a means of doing so. We demonstrate a methodology that allows a person to interact “in the flesh” with a conversational agent whose interface is an actual human body.

Contemporary android science

Android science aims to develop artificial systems identical to humans in both appearance and behavior (verbal and non-verbal) for the purposes of exploring human nature and investigating the ways in which these systems might integrate into human society (MacDorman and Ishiguro, 2006a; Ishiguro and Nishio, 2007). The field is as interested in better understanding people through their interacting with anthropomorphic technology as it is in further developing the technology itself. Considerable progress has been made in these endeavors, with perhaps the most notable work being that undertaken and inspired by Hiroshi Ishiguro of Osaka University’s Intelligent Robotics Laboratory, whose research and engineering teams have developed highly lifelike autonomous and semi-autonomous androids. MacDorman and Ishiguro (2006b) argue that in being controllable, programmable, and replicable, androids are in certain respects superior to human actors as social and cognitive experimental stimuli. They further contend that androids can evoke in humans expectations and emotions that attenuate the psychological barrier between people and

machines.

The motor behaviors of autonomous androids are controlled by technologies that perceive and orient to the physical environment while their speech is controlled by a conversational agent. As autonomous technologies are still quite limited in terms of functionality, the social capacities of these types of androids are severely constrained. Tele-operated androids, meanwhile, overcome the limitations of fully autonomous models by-way-of a human operator controlling the android's speech and movement (Nishio et al., 2007b). On account of their enhanced social capabilities, tele-operated androids have stimulated ample research in psychology and other domains of social and cognitive science. For instance, researchers have investigated the extent to which a person's presence with remote others is amplified or weakened when tele-operating an android compared to when communicating in person or via more distal technological mediators such as video conferencing (Nishio et al., 2007a; Sakamoto et al., 2007). Researchers have also explored the extent to which tele-operators perceive their android to be extensions of themselves, sensing physical stimuli administered to the android as if the stimuli had been administered to their own body (Ogawa et al., 2012). Perhaps the most discussed phenomenon in the field of android science is the "uncanny valley," posited by Mori (1970). This idea suggests that the affinity a person has for an artificial agent will increase as the appearance and motor behavior of the agent becomes more human-like; however, at a certain point along the human-likeness continuum (where the agent begins to look more or less human but for slight, yet telling, signs of artificiality) feelings of affinity will sharply decline, before rapidly rising again as the agent becomes indistinguishable from an actual human (MacDorman and Ishiguro, 2006b; Seyama and Nagayama, 2007).

We propose inverting the composition of tele-operated android systems in order to create hybrid entities consisting of a human whose words (and potentially motor actions) are entirely or partially determined by a computer program. We refer to such hybrids as "echoborgs," which can be classified as a type of "cyranoïd"—Milgram's (2010) term for a hybrid composed of a person who speaks the words of a separate person in real-time. Echoborgs can be used to examine the role of the human body, as the delivery mechanism of communication, in mediating social emotions, attributions, and other interpersonal phenomena emergent in face-to-face interaction. Furthermore, echoborgs can be used to evaluate the performance and perception of artificial conversational agents under conditions

wherein people assume they are interacting with an autonomously communicating human being. To ground these claims, however, we shall first discuss the tools and constraints of contemporary android science in order to identify where echoborg methodology can contribute.

The challenge of creating androids that speak autonomously. Examples of autonomous androids include Repliee Q1 and Repliee Q2, which were developed jointly by Osaka University and the Kokoro Corporation (see Ishiguro, 2005; Ranky and Ranky, 2005). Because androids of this nature attempt to replicate humans at both an outer/physical level as well as an inner/dispositional level, they can be evaluated against what Harnad (1991) defined as the Total Turing Test (also referred to as the Robotic Turing Test; Harnad, 2000), which establishes the entire repertoire of human linguistic and sensorimotor abilities as the appropriate criteria for judging machine imitations of human intelligence. The development of an autonomous android capable of passing such a test, however, remains a distant holy grail.

One source of current constraints concerns how artificial agents in general interpret and participate in dialog. Various terminologies describe technology that interacts with humans via natural language. “Dialog system,” “conversational agent,” and “conversational AI,” for instance, are terms used to denote the linguistic subsystems of artificial agents, though no clear consensus exists with regard to how non-overlapping these and other terms are. “Conversational agent,” the term we have employed thus far, is perhaps the most convenient term for conceptualizing the echoborg because it has been adopted by a parallel project—the development of embodied conversational agents (software that interfaces through onscreen anthropomorphic avatars). Much of the literature that distinguishes the functionality of various linguistic subsystems, however, couches these technologies as dialog systems. Types of dialog systems include high-level systems of integrated artificial intelligence that employ advanced learning and reasoning algorithms enabling a user and a machine to jointly accomplish specific tasks within a formal dialog structure (e.g., logistics and navigation planning agents), low-level systems that use basic algorithms to simply mimic, rather than understand, casual human conversation (e.g., web-based “chat bots”), and mid-level systems that strike a balance between high-level and low-level functionality (e.g., agents designed to field queries from and respond to pedestrians in transit centers; for a discussion of dialog system hierarchy, see Schumaker et al., 2007). Dialog systems can also be differentiated in

terms of the level of initiative they take when interacting with users (Zue and Glass, 2000). System-initiative agents are those that control the parameters of dialog and elicit information from the user that must be compatible with certain response formats (e.g., interactive voice response telephone systems). User-initiative agents, on the other hand, are those in which the user presents queries to a passive agent (e.g., Apple’s Siri application). Mixed-initiative agents (by far the least developed variety; Mavridis, 2015) involve both the user and agent taking active roles in a joint task with the nature of dialog being qualitatively more conversational relative to other types of dialog systems.

If we treat, as Turing (1950) did, discourse capacity as a basic proxy for an interlocutor’s “mind,” then even today’s most advanced dialog system technologies render available to artificial agents such as androids minds that are at best starkly non-human (though potentially very powerful), and at worst extremely impoverished relative to that of humans. Though contemporary high-level and mid-level dialog systems are indeed impressive and their functionality continues to expand rapidly, they are not, in principle, attempts to mimic a human interlocutor capable of casual conversation. On the contrary, they are presently intended to interact with humans in specific domains and generally do not operate outside of these contexts (e.g., such a system cannot spontaneously switch from being a logistics planning agent to having a conversation about an ongoing basketball game). No human would be expected to communicate in a manner similar to these types of artificial intelligence, nor are humans necessarily constrained in terms of only being capable of communicating from within a fixed and narrow language-game. System-initiative and user-initiative agents also deviate from the norms of human–human interaction as they grant to one interlocutor total and unbreakable communicative control.

Though we can perhaps imagine high-level and mid-level dialog systems capable of engaging humans in casual conversation someday being ubiquitous throughout social robotics, at present only certain low-level and primarily text-based systems are engineered specifically for this purpose. An early but well-known example of such a system is ELIZA, a chat bot with the persona of a Rogerian psychotherapist (Weizenbaum, 1966). Modern examples include A.L.I.C.E. (Artificial Linguistic Internet Chat Entity; Wallace, 2015), Cleverbot (Carpenter, 2015), Mitsuku (Worswick, 2015), and Rose (Wilcox, 2015). Many chat bots make use of the highly customizable AIML (Artificial Intelligence Markup Language) XML dialect developed by Wallace (2008) and operate by recognizing word patterns delivered by a

user and matching them to response templates defined by the bot's programmer. Increasingly sophisticated mechanisms for generating response corpora have been developed for chat bots in recent years. For instance, some developers have turned to real-time crowdsourcing of online communication repositories, such as Twitter and Facebook, as a means of producing responses appropriate for a given user input (see Mavridis et al., 2010; Bessho et al., 2012).

Chat bots are widely available on the internet and feature regularly in events such as the annual Loebner Prize competition (Loebner, 2008), a contest held to determine which chat bot performs most successfully on a Turing Test. This test involves a human interrogator simultaneously communicating via text with two hidden interlocutors while attempting to uncover which of the two is a bot and which is a real person. To date, no chat bot has reliably passed as a human being, and we are unlikely to see this feat accomplished in the near future (Dennett, 2004; French, 2012).

Generally, human interactions with chat bots fail to arrive at what conversation analysts refer to as "anchor points": mutually attended to topics of shared focus that establish an implicit "center of gravity" during moments of conversation following routine canonical openings (Schegloff, 1986; Friesen, 2009). As chat bots tend to be user-initiative agents, they cannot engage in the type of fluid mixed-initiative conversation that is natural to mundane human-human interaction (Mavridis, 2015). Chat bots demonstrate a poor capacity to reason about conversation, cannot consistently identify and repair misunderstandings, and generally talk at an entirely superficial level (Perlis et al., 1998; Shahri and Perlis, 2008). According to Raine (2009), many chat bots work "based on an assumption that the basic components of a communication are on a phrase-by-phrase basis and that the most immediate input will be the most relevant stimulus for the upcoming output" (p. 399), an operative model that can lead conversation to irreparably fall apart when the perspectives of parties to a conversation diverge in terms of the meaning or intention each party assigns to an utterance. Human communication is fundamentally temporal and sequential, with many past and possible future utterances feeding into the meaning of a given utterance (Linell, 2009).

Developing acoustic technology that can accurately perceive spoken discourse remains a related challenge. The error rate of speech recognition technology is dramatically compounded by, among other things, variation in a speaker's accent, the lengthiness and spontaneity of their speech, their use of contextually specific vocabulary, the presence of

multiple and overlapping speakers, speech speed, and so on (Pieraccini, 2012). Thus, speech recognition systems within artificial agents perform best not when discerning casual conversational dialog, but when discerning brief and predictable utterances. Microphone array technologies and software capable of identifying and isolating multiple speakers continue to improve (e.g., the “HARK” robot audition system; Nakadai et al., 2010; Mizumoto et al., 2011), but demonstrations of these systems have essentially involved stationary apparatuses confined to laboratory environments.

Tele-operated androids: Mechanical bodies, human operators. Tele-operated androids were developed in part to overcome a social research bottleneck within android science born of the various limitations of conversational agents and perception technologies (Nishio et al., 2007b; Watanabe et al., 2014). They thus constitute a methodological trade-off: rather than being both physically artificial and having computer-controlled behavior (a combination that currently results in poor social functioning), the tele-operated paradigm cedes behavioral control to a human, thus augmenting the speech and motor capabilities of the android.

Perhaps the most well-known tele-operated android is Geminoid HI-1, a robot modeled in the likeness of its creator, Hiroshi Ishiguro. From a remote console, the tele-operator is able to transmit their voice through the geminoid (derived from the Latin word “geminus,” meaning “double”) while software analyzing video footage of the tele-operator’s body and lip movements replicate this motor behavior in the geminoid. The tele-operator can also manually control specified behaviors such as nodding and gaze-direction. Video monitors and microphones capture the audio-visual perspective of the geminoid and transmit to the tele-operation console, allowing the tele-operator to observe the geminoid’s social environment (Nishio et al., 2007b; Becker-Asano et al., 2010).

Relative to their fully-autonomous counterparts, the enhanced conversational capacities of tele-operated androids allow researchers to study communicatively rich human–android interactions as well as offer a means of operationally separating the behavioral control unit of an agent (the tele-operator) from the body, or interface, of the agent (the android). As Nishio et al. (2007b) contend:

The strength of connection, or what kind of information is transmitted between the body and mind, can be easily

reconfigured. This is especially important when taking a top-down approach that adds/deletes elements from a person to discover the “critical elements” that comprise human characteristics” (p. 347).

These methodological assets have inspired an abundance of exploratory laboratory and field work in recent years. Abildgaard and Scharfe (2012), for instance, used Geminoid-DK to conduct university lectures and reported on how perceptions of the android differed between male and female students. Research involving android-mediated conversations between parents and children has explored to what extent children sense the personal presence of a tele-operator (Nishio et al., 2008). Straub et al. (2010) studied how tele-operators and those they communicate with jointly construct the social identity of an android. Dougherty and Scharfe (2011), meanwhile, explored whether touch influences a person’s trust in a tele-operated android.

Despite the progress and promise of tele-operated androids, this line of research faces particular constraints. The non-verbal behaviors of autonomous and semi-autonomous androids are more mechanical and less fluid relative to humans. In their neuroimaging analysis of how people perceive geminoid movement, Saygin et al. (2012) show how incongruity between appearance (human-like) and motion (non-human-like) implicitly violates people’s expectations. Developing tools for matching an android’s bodily movements to those of its tele-operator is a major research priority (Nishio et al., 2007b), and improving techniques for achieving facial synchrony is particularly necessary given the intricate facial musculature of humans and the role of facial expression in conveying emotion and facilitating social interaction (Ekman, 1992; Bänziger et al., 2009; for a discussion of robot emotion conveyance, see Nitsch and Popp, 2014). Current anthropomorphic androids are relatively limited in terms of their capacity for human-like facial expressivity (Becker-Asano, 2011). For instance, Geminoid F’s face can successfully express the emotions sad, happy, and neutral, but the model struggles to convincingly convey angry, surprised, and fearful (Becker-Asano and Ishiguro, 2011). Also, the inexactness of an android’s lip movements in relation to the words spoken by its tele-operator has been discussed as possibly degrading the quality of social interactions (Abildgaard and Scharfe, 2012). Moreover, geminoids and other android models cannot walk on account of their having large air compressors facilitating numerous pneumatic actuators (Ishiguro and Nishio, 2007).

The imperfect appearance of tele-operated androids remains a barrier to replicating the social psychological conditions of face-to-face human–human interaction. Despite painstaking efforts to create realistic silicone android models (Ishiguro and Nishio, 2007), people are minutely attuned to subtle deviations from true humanness (e.g., eyes that lack glossy wetness). In a field study conducted to test whether people would notice an inactive or relatively passive geminoid in a social space, a majority of people reported having seen a robot in their surroundings (von der Pütten et al., 2011), a finding which suggests that most people are not easily fooled into believing an android is an actual person even in social situations where they do not engage the android directly. Moreover, though geminoids and other highly anthropomorphic androids are seen as the most human-like and least unfamiliar of robot types, people nonetheless perceive these androids as more threatening than less anthropomorphic models (Rosenthal-von der Pütten and Krämer, 2014).

There is also an important practical constraint characterizing the tele-operated and autonomous android paradigms. As Ziemke and Lindblom (2006) point out, it is quite time consuming and costly to produce android experimental apparatuses. This raises issues as to the scalability of the current android science research model and the extent to which experiments making use of a particular device in one laboratory can be replicated elsewhere.

The echoborg

An echoborg is composed of a human whose words (and potentially motor actions) are entirely or partially determined by a computer program. Echoborgs constitute a methodological trade-off inverse to that of the tele-operated paradigm discussed above, as they allow the possibility of studying social interactions with artificial agents that have truly human interfaces. The unique affordances of echoborgs can complement those of tele-operated and fully-autonomous androids and contribute to our understanding of the social psychological dynamics of human–agent interaction.

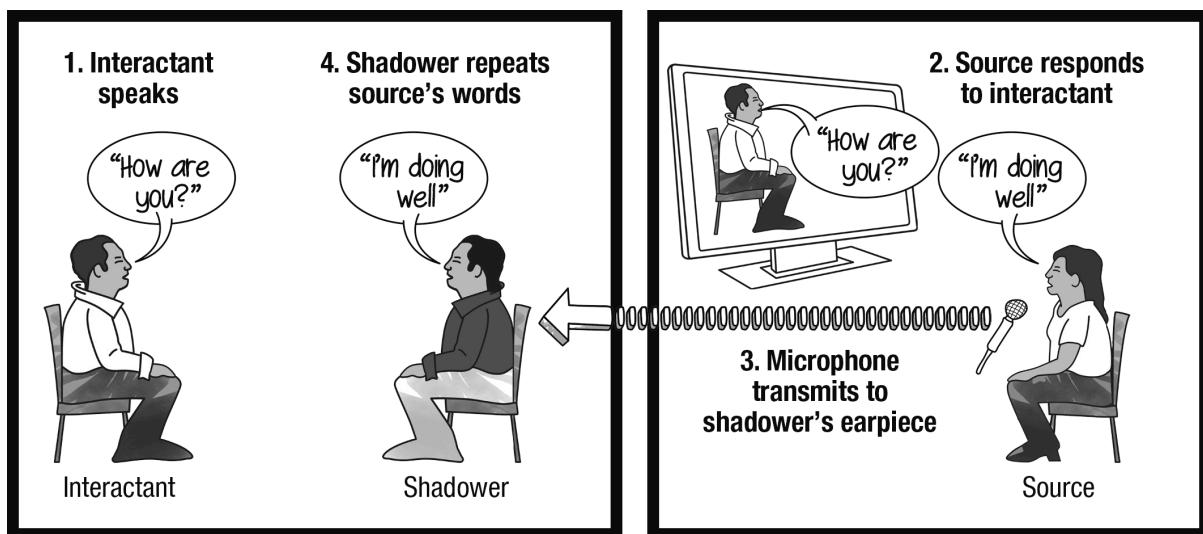
Speech shadowing and the cyranoid method. The echoborg concept stems from work conducted by Corti and Gillespie (2015), whose application of Milgram’s (2010) “cyranoid method” of social interaction demonstrates a means of creating hybrid human entities via an audio-vocal technique known as “speech shadowing.” Speech shadowing involves a person (the shadower) voicing the words of an external source simultaneously as those words are heard (Schwitzgebel and Taylor, 1980). This can be facilitated by-way-of an inner-ear

monitor worn by the shadower that receives audio from the source. Research has shown that native-language shadowers can repeat the words of a source at latencies as low as a few hundred milliseconds (Marslen-Wilson, 1973, 1985; Bailly, 2003) and can perform the technique while simultaneously attending to other tasks (Spence and Read, 2003). Shadowers tend to reflexively imitate certain gestural elements of their source (e.g., stress, accent, and so on)—a phenomenon known as “phonetic convergence” (Goldinger, 1998; Shockley et al., 2004; Pardo et al., 2013).

One finds the use of speech shadowing as a research tool primarily in psycholinguistics and the study of second-language acquisition. In the late 1970s, however, Milgram—famous for his controversial studies on obedience to authority (Milgram, 1974)—began using speech shadowing to investigate social scenarios involving people communicating through shadowers. He saw the technique as a means of pairing sources and shadowers whose identities differed in terms of race, age, gender, and so on, thus allowing sources to directly experience an interaction in which their outer appearance was markedly transformed (see Figure 3.1). From the point of view of the shadower, the method enabled exploration into the sensation of contributing to an unscripted conversation not one’s self-authored thoughts, but entirely those of a remote source. Inspired by the play *Cyrano de Bergerac*, the story of a poet (Cyrano) who assists a handsome but inarticulate nobleman (Christian) in wooing a woman by telling him what to say to her, Milgram referred to these source-shadower pairs as “cyranoids.”

As speech shadowing proved to be a relatively simple task that research participants were quick to grasp, Milgram quickly began exploring a variety of cyranic interactions. For instance, in several pilot studies he examined whether “interactants” (Milgram’s term for those who encountered a cyranoid) would notice if the source was changed mid-conversation (Milgram, 1977). Milgram (2010) also sourced for 11- and 12-year-old children during interviews with teachers naïve to the manipulation. Following these interactions, all of the teachers seemed to take the interviews at face value—they neither picked up on the true nature of the interactions nor sensed that the child they interviewed had behaved non-autonomously. The teachers had succumbed to the “cyranic illusion,” that is, the tendency to perceive interlocutors as autonomous communicators and thus fail to notice an interlocutor that is a cyranoid.

Figure 3.1: *Illustration of a basic cyranoid interaction*. The shadower voices words provided by the source while engaging with the interactant in person.



Corti and Gillespie (2015) argue that one of the cyranoid method's primary strengths is that it allows the researcher to manipulate one component of the cyranoid, either the shadower or the source, while keeping the other component fixed. Thus, one can study how the same source is perceived when interacting through a variety of shadower-types. Conversely, a researcher can opt to keep the shadower constant and vary the identity of the source across experimental conditions. This capacity mirrors the functionality of tele-operated androids as well as similar methods for studying transformed social interactions (e.g., using 3D immersive virtual environment technology to alter people's identities; see Blascovich et al., 2002; Bailenson et al., 2005; Yee and Bailenson, 2007). A unique benefit of the cyranoid method is that it allows for in person, face-to-face interactions between an interactant and a hybrid. When interacting with a cyranoid, one is not interacting with an onscreen person, or a human-like machine, or a virtual representation of a human, but with an actual human body.

While Corti and Gillespie's (2015) recent work was conducted in the laboratory, it follows recent field explorations of cyranoids in experiential art installations (Mitchell, 2009) and as classroom learning tools (Raudaskoski and Mitchell, 2013). Taken together, these studies outline a number of basic protocols for constructing cyranic interactions and discuss the devices necessary for creating a basic cyranoid apparatus, which involves both a means of discreetly transmitting audio from the source to the shadower as well as a means for the

source to hear (and, if possible, see) the interaction between the shadower and the interactant. The amalgam of devices one uses toward these requirements depends upon the type of interaction the researcher wishes to create. For instance, if a researcher wants to keep hidden from interactants the fact that a cyranoid is present in an interaction, then the cyranoid apparatus should be discreet and non-visible/audible to interactants. If the researcher wants the shadower to be mobile, then the devices that compose the cyranoid apparatus must transmit wirelessly. Minimizing the audio latency in the communication loop is crucial to any cyranoid apparatus; interactant→source and source→shadower audio transfer must be accomplished in a realistic amount of time.

A cyanic interaction involving a covert cyranoid is typically accomplished using an apparatus similar to the following. A wireless “bug” microphone placed near where the shadower and interactant engage each other transmits to a radio receiver listened to by the source in an adjacent soundproof room. The source speaks into a microphone connected to a short-range radio transmitter which relays to a receiver worn in the pocket of the shadower. Connected to the shadower’s receiver is a neck-loop induction coil worn underneath their clothing. The shadower wears a wireless, flesh-colored inner-ear monitor that sits in their ear canal and receives the signal emanating from the induction coil, allowing the shadower to hear and thus voice the source’s speech. This amalgam of devices is neither visible nor audible to interactants.

Ceding verbal agency to a machine. Echoborg methodology takes the original cyranoid model and replaces the human source with an artificial conversational agent. The words produced by the conversational agent are thus voiced and embodied by a human shadower. Echoborgs have at least four main research affordances:

Interchangeability of shadowers and conversational agents. Both the shadower and the conversational agent that comprise an echoborg are easily customizable and interchangeable. The researcher need only train a confederate with the desired physical attributes to speech shadow sufficiently and then couple them with a conversational agent. This gives the researcher the freedom to construct many echoborgs, each differentiated from one another in terms their particular conversational agent, gender, age, and so on. Thus, one can observe how the same conversational agent is perceived depending on the identity of the shadower by holding the conversational agent constant across experimental conditions and varying the

shadower (e.g., female shadower vs. male shadower). Alternatively, the researcher can hold the shadower constant and vary the conversational agent (e.g., ELIZA vs. A.L.I.C.E).

Visual realism. Echoborgs offer a means of studying interactions under conditions where the interactant's cognitive sense of the interaction is undistorted by any esthetic, acoustic, non-verbal, or motor non-humanness of the physical agent they encounter (e.g., lips that do not exactly align with the words they utter or eyes that do not perfectly make contact with the interactant's). Speech shadowing is not a cognitively demanding task; it is rather simple for a well-rehearsed speech shadower to attend to other behaviors while replicating the speech of their source, including matching their body language to the words they find themselves repeating (e.g., shaking their head from side-to-side upon articulating the word “no”).

Mobility. Echoborgs can take advantage of the shadower's physical mobility and need not be confined to stationary interactions—they can walk or otherwise move about while communicating with interactants. Human communication did not evolve for having conversations *per se*; it evolved for coordinating joint activity (Tomasello, 2008). Research on everyday language use shows that communication is a means of doing (Clark, 1996). Accordingly, mobile echoborgs open up the possibility of testing conversational agents in the context of performing a joint non-stationary activity.

Covert capacity. Taking advantage of the cyranic illusion, echoborgs can interact with people covertly (i.e., under conditions wherein interactants assume they are encountering an autonomously communicating person). This affordance can be juxtaposed with the fact that at present, those who interact with tele-operated or autonomous androids are under no illusion that they are interacting with a fully-autonomous human being. The covert capacity of echoborgs thus presents a new means of researching interactions with conversational agents. It is one thing to evaluate interactions with conversational agents in contexts where people are cognitively aware, or at least primed to believe, that they are speaking to something artificial, but it is entirely different to study these systems under conditions where the interface one encounters (an actual human body) creates the visceral impression that one is dealing with an autonomous person.

Overview of studies

We conducted three experiments in which participants interacted with echoborgs. These

studies explored the ways in which echoborgs, as human interfaces, mediate the experience of conversing with a chat bot in various contexts, as well as the extent to which echoborgs improve a chat bot’s ability to pass as human (i.e., be taken for a human rather than a robot). Each study was approved by an ethics review board at the London School of Economics and Political Science and conducted at the university’s Behavioral Research Laboratory. Adult participants were recruited online via the university’s research participant recruitment portal and included students from the university, university employees, and people unaffiliated with the university. Participants gave informed consent prior to participation and were debriefed extensively.

Study 1: Turing testing with echoborgs

Aims. In outlining the logic of his imitation game, Turing (1950) argued that “there was little point in trying to make a “thinking machine” more human by dressing it up in such artificial flesh” (p. 434) and made a clear distinction between what he thought of as the physical (likeness) and intellectual (functional) capacities of humans. However, this distinction has been criticized (Harnad, 2000); perceiving the salient bodily characteristics of other entities is fundamental to how humans infer the subjective states (or lack thereof) of said entities, be they real or unreal in reality (Graziano, 2013). To explore this tension, our first study investigated a Turing Test scenario wherein participants were asked to determine which of two shadowed interlocutors was truly human and which was a chat bot. Furthermore, we sought to determine whether a chat bot voiced by a human shadower would be perceived as more human-like than the same bot communicating via text.

Shadowers and subjects. Two female graduate students (both aged 23) were trained as speech shadowers. Eighty-two participants (42 female, mean age = 28.93, $SD = 12.05$) were randomly assigned into pairs within one of two experimental conditions: Text Interface ($n = 21$) and Echoborg ($n = 20$). One participant within each pair was randomly selected to function as the Turing Test interrogator while the second participant was designated as the human interlocutor. In all pairs, participants were both unfamiliar with one another and unaware of the other’s role in the study.

Procedure. From the interaction room, the researcher instructed the interrogator that the study involved using a text-based instant messaging client (Pidgin) to simultaneously communicate with two anonymous interlocutors, one of whom was a chat bot (Cleverbot).

The interrogator's computer showed two separate text-input windows, one that delivered to "Interlocutor A," and another that delivered to "Interlocutor B." The interrogator was told that following 10-min of conversation they would be asked which of these two interlocutors they believed was the real human. Meanwhile, in a separate room, a research assistant instructed the human interlocutor that the study involved holding a 10-min conversation with a stranger and that their task was to simply respond to messages that appeared on a computer screen. The human interlocutor was thus blind to the fact that they were engaged in a Turing Test. Both the interrogator and the human interlocutor were informed that they were free to discuss any topic during the interaction so long as nothing was vulgar.

Text interface condition. Once instruction was complete, the researcher relocated to a third room (the source room) where they monitored the interaction using a computer. Messages that the interrogator typed to Interlocutor A were routed to the researcher, who input the received text into Cleverbot and routed Cleverbot's response back through the instant messaging client to the interrogator. Messages the interrogator sent to Interlocutor B, meanwhile, were routed to the human interlocutor's computer, and the human interlocutor directly responded in text via the instant messaging client.

Echoborg condition. The interrogator was further instructed that though they would type messages to Interlocutor A and Interlocutor B via the instant messaging client, the responses of these two interlocutors would be spoken aloud by two speech shadowers. The two speech shadowers, with shadowing equipment, entered the room, sat side-by-side facing the interrogator at a distance of roughly six feet, and it was made known to the interrogator which shadower would reproduce the words of Interlocutor A and which would reproduce the words of Interlocutor B (shadowers alternated between trials in terms of the interlocutor they were paired to). The interrogator was informed that the shadowers would speak solely words they received from their respective sources and that at no point during the interaction would the shadowers speak self-authored thoughts. Furthermore, the interrogator was informed that both interlocutors would only respond to typed messages and that nothing the interrogator spoke aloud would be responded to.

Following these instructions, the researcher relocated to the source room. As in the Text Interface condition, messages that the interrogator sent to Interlocutor A were routed to the researcher's computer where they were input by the researcher into Cleverbot. Instead of

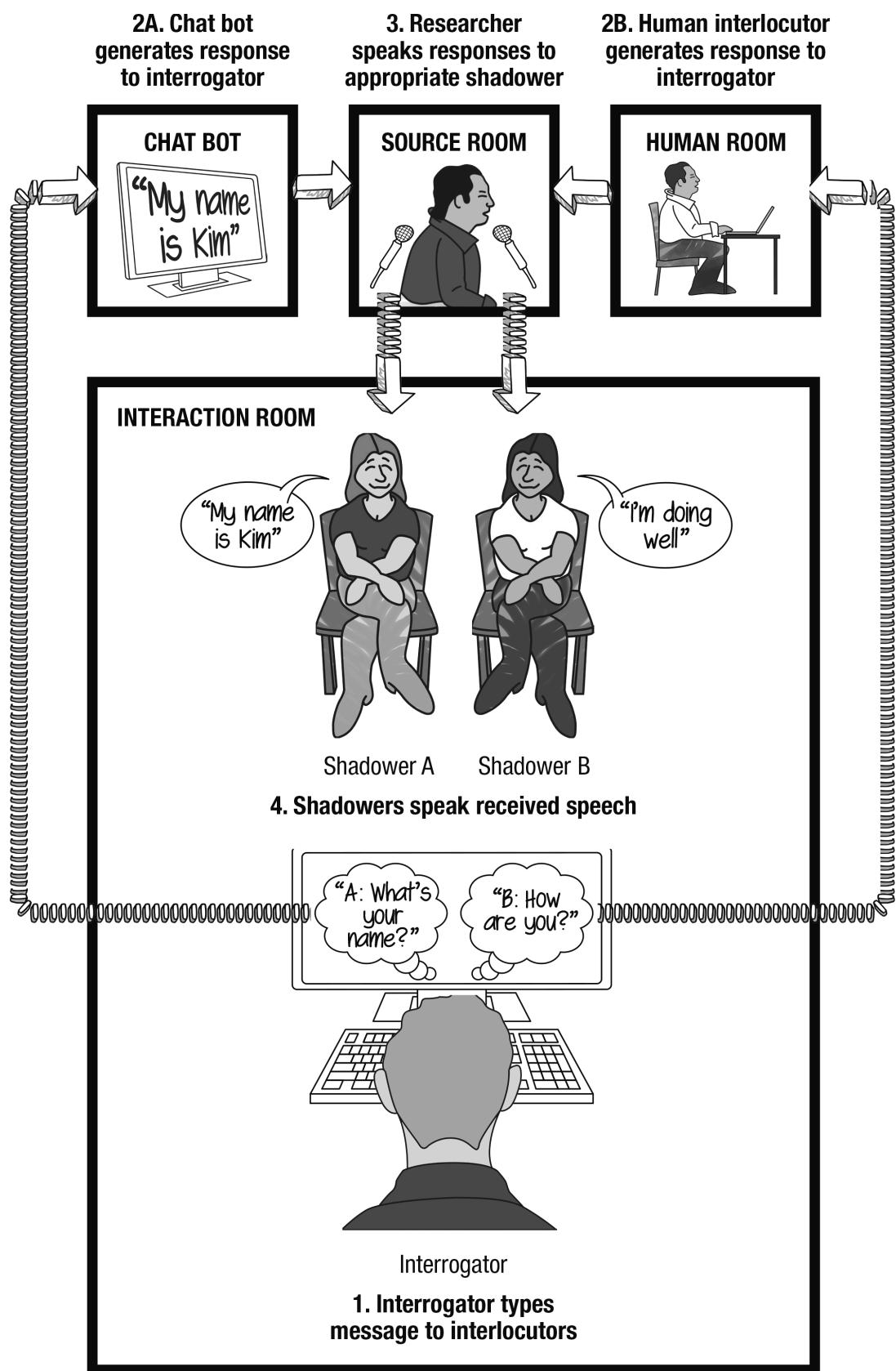
routing Cleverbot's responses back to the interrogator through the instant messaging client, however, the researcher spoke Cleverbot's responses into a microphone which relayed to the speech shadower paired to Interlocutor A, thus allowing them to hear and repeat Cleverbot's words to the interrogator. Similarly, the human interlocutor's typed responses were routed to the researcher's computer (rather than directly to the interrogator), allowing the researcher to speak these messages into a separate microphone which relayed to the shadower paired to Interlocutor B (see Figure 3.2).

Stock responses. Cleverbot's response formats are not programmed; Cleverbot references past conversations it has held with people over the internet when generating a reply to a given user input (Carpenter, 2015). Unlike other bots, therefore, Cleverbot has no consistent identity. Its strength lies in its ability to learn unique ways of responding. We decided, however, that in order to establish consistency between experimental trials, three stock responses would be supplied in both conditions to the interrogator in lieu of a response generated by Cleverbot. Each time the interrogator inquired as to the name of Interlocutor A, the standard response "My name is Kim" was supplied to the interrogator. In response to questions as to what Interlocutor A's occupation was, the response "I'm a psychology student here" was supplied. Finally, in response to questions concerning where Interlocutor A was from, the response "I'm from London" was given.

Measures. Following the interaction, the interrogator indicated on a questionnaire which of the two interlocutors (A or B) they believed was the real human and indicated along a 10-point scale how confident they were that they had made the correct identification (1: not at all confident; 10: highly confident). Interrogators also rated each interlocutor along a 10-point scale in terms of how human-like they seemed (1: seemed very mechanical and computer-like; 10: seemed very human-like).

Results. In the Text Interface condition, 21 out of 21 interrogators correctly identified Interlocutor B as being the real human, compared to 18 out of 20 interrogators in the Echoborg condition, a non-significant difference, $z = 1.49, p = 0.14$ (two-tailed). There was no significant difference between conditions in terms of how confident interrogators were with regard to their answers, with interrogators in the Text Interface condition reporting an average confidence of 7.67 ($SD = 2.61$) and interrogators in the Echoborg condition reporting an average confidence of 7.55 ($SD = 1.70$), $t(39) = 1.68, SE = 0.69, p = 0.87$.

Figure 3.2: Illustration of a Turing Test scenario involving speech shadowing. This figure visually depicts the Echoborg condition in Study 1.



Human-likeness ratings were compared using a repeated measures analysis of variance, Condition (Text Interface vs. Echoborg) treated as a between-subjects factor and Interlocutor (Interlocutor A vs. Interlocutor B) treated as a within-subjects factor. There was a significant main effect of Interlocutor showing that Interlocutor B was perceived as significantly more human-like than Interlocutor A in both conditions, $F(1,39) = 130.87, r = 0.88, p < 0.001$. There was also a significant interaction between Condition and Interlocutor, $F(1,39) = 7.23, r = 0.40, p < 0.05$. Independent samples means tests showed that the average human-likeness rating of Interlocutor A in the Text Interface condition ($M = 2.14, SD = 1.15$) was significantly less than the average rating in the Echoborg condition ($M = 4.05, SD = 2.42$), $t(39) = -3.25, SE = 0.59, p < 0.01$. Meanwhile, the average human-likeness rating of Interlocutor B in the Text Interface condition ($M = 8.76, SD = 1.51$) was not significantly different from the average rating in the Echoborg condition ($M = 8.15, SD = 1.46$), $t(39) = 1.32, SE = 0.46, p = 0.20$.

Discussion. The interface (human body vs. text) engaged by the interrogator made no statistically significant difference in terms of their ability to discern which interlocutor was the real human. The chat bot, however, was perceived by interrogators as significantly more human-like when being shadowed by a person compared to when simply communicating via text. This contrasted with the fact that how human-like human interlocutors seemed to participants did not depend on whether their words were voiced by a speech shadower. This suggests that as the quality of an interlocutor's discourse capacity improves (i.e., becomes more human) in Turing Test scenarios, the role the interface plays in eliciting judgments about human-likeness declines.

Study 2: A human imitating a chat bot?

Aims. Study 2 investigated whether attributing human agency to an interlocutor is increasingly determined by the nature of the interface as the words spoken by the interlocutor provide less definitive evidence. We designed a scenario wherein participants encountered an interlocutor and had to determine whether the interlocutor was (a) a person communicating words that had been generated by a chat bot, or (b) a person merely imitating a chat bot, but nonetheless speaking self-authored words (the former option always being true). The point here was to see whether or not the interface participants encountered (human body vs. text) influenced whether they thought their interlocutor was producing self-authored words or, alternatively, those of a machine. The framing of the scenario leads participants to expect that

the communication offered by their interlocutor will be abnormal, thus the conversational limitations of chat bots are not a liability as they are in standard Turing Test scenarios. By design, participants must form an attribution regarding the communicative agency of their interlocutor under conditions of ambiguity.

Research on perceptual salience suggests that people will deem causal what is salient to them in the absence of equally salient alternative explanations (Jones and Nisbett, 1972; Taylor and Fiske, 1975). Dual process information evaluation theories propose that when a person evaluates the communication and behavior of others, stimulus ambiguity increases reliance on heuristic cues (e.g., appearance) at the expense of more thoughtful situational evaluation (Sager and Schofield, 1980; Devine, 1989; Chen and Chaiken, 1999). We extrapolated from this research that when faced with an ambiguous situation in which one's interlocutor was either truly speaking words generated by a chat bot or merely pretending to be one, the interface (and thereby the heuristic cues) salient to the participant would determine how they attributed authorship to the words they encountered. We therefore hypothesized that those who encountered an echoborg would be more likely to see their interlocutor as producing self-authored words (imitating a chat bot) compared to those who encountered an interlocutor through a text interface.

Shadowers and subjects. A female graduate student (aged 30) was trained to perform as a speech shadower. Fifty-eight adult participants (35 female; mean age = 25.19, $SD = 9.08$) were randomly assigned to one of two conditions: Echoborg ($n = 28$) and Text Interface ($n = 30$).

Procedure. As with Study 1, Cleverbot, as well as the three stock responses described above, were used in all trials.

The participant was led to an interaction room and instructed by the researcher that the study involved holding a 10-min conversation with an interlocutor who was either (a) communicating solely words that had been generated by a chat bot program (at no point speaking anything self-authored), or (b) simply imitating a chat bot program, but producing self-authored words nonetheless. The researcher ensured that the distinction between these scenarios was clear to the participant and gave the further instruction that the participant would be asked following the interaction which of the two scenarios they believed to have

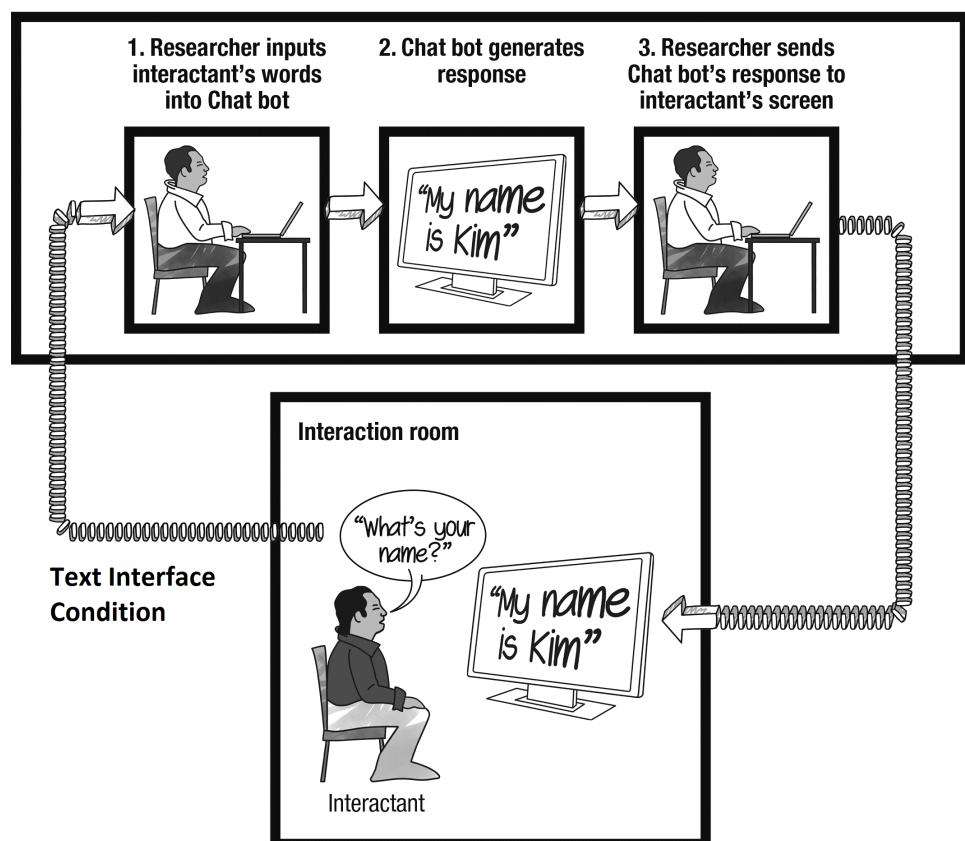
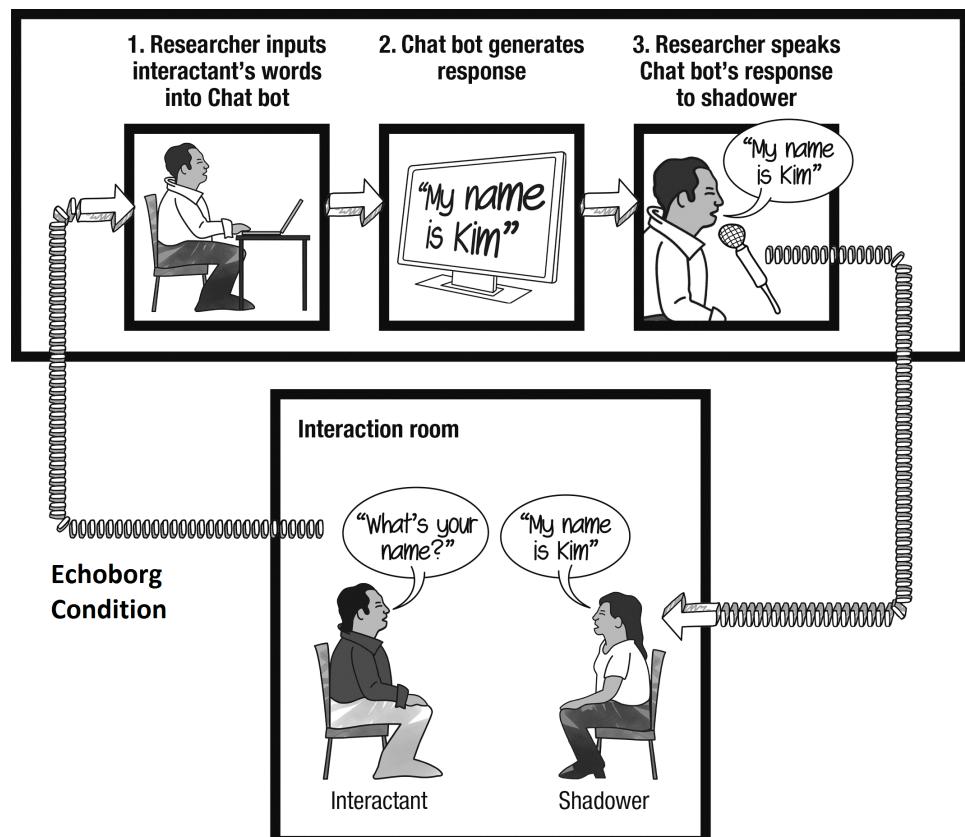
been the case. The participant was informed that they were free to discuss anything they liked with their interlocutor so long they refrained from vulgarity.

Unlike Study 1, which had participants send messages to their interlocutors via an instant messaging client, Study 2 featured participants speaking aloud to their interlocutor as they would during any other face-to-face encounter, thereby increasing the mundane realism of the scenario. The apparatus for this type of interaction, however, required a means of inputting the participant's spoken words into the chat bot in the form of text. As we deemed speech-to-text software to be insufficient for our purposes (being too slow and inaccurate), we settled on a procedure wherein the researcher (from an adjacent room) acted as the chat bot's ears and speed typed the participant's words into the chat bot as they were being spoken, paraphrasing when necessary for particularly verbose turns. This can be conceptualized as a minimal technological dependency format of the echoborg method (as opposed to a full technological dependency format which would place acoustic perception solely on technology). Although a minimal technological dependency format adds an additional human element to the communication loop, it ensures that accurate representations of interactants' words are processed by the conversational agent.

Text interface condition. The participant was seated in front of a computer screen which displayed a blank instant messaging client chat window. The participant was instructed that they were to address their interlocutor by speaking aloud and that their interlocutor would respond via text readable in the chat window. Once instruction was complete, the researcher left the interaction room and returned to the adjacent source room. From the source room, the researcher overheard words spoken by the participant via a covert wireless microphone and speed typed them into Cleverbot's text-input window. Cleverbot's responses were then sent through the instant messaging client to the participant's screen in the interaction room (see Figure 3.3).

Echoborg condition. The participant was instructed that as soon as the researcher left the interaction room their interlocutor would enter and sit facing the participant (at a distance of roughly six feet). The participant was not made aware of the fact that their interlocutor would be wearing an earpiece and receiving messages via radio, and the cyranoid apparatus was not visible to the participant. The researcher then left the interaction room and returned to the adjacent source room while the shadower entered the interaction room and sat across from the participant. The researcher listened to the words of the participant via a covert wireless

Figure 3.3: Illustration of interaction scenarios in Study 2 and Study 3.



microphone, speed typed them into Cleverbot's text-input window, and subsequently spoke Cleverbot's responses into a microphone which relayed to the shadower's inner-ear monitor.

Measures. Following the interaction, the participant indicated on a questionnaire whether they thought their interlocutor had truly been producing words generated by a chat bot program or whether their interlocutor was simply imitating a chat bot.

Results. Of the 30 participants in the Text Interface condition, 11 stated following the interaction that they believed their interlocutor was simply imitating a chat bot compared to 22 of 28 participants in the Echoborg condition. A binary logistic regression model showed these proportions to be significantly different from one another, $OR = 6.33$, $b = 1.85$, $SE = 0.60$, $p < 0.01$ (indicating that the odds of a participant in the Echoborg condition deciding their interlocutor was imitating a chat bot were 6.33 times greater than the odds of a participant in the Text Interface condition coming to the same conclusion).

To gain a sense of the audio latency dynamics of echoborg interactions involving minimal technological dependency, we randomly selected four trials from the Echoborg condition and measured the time between the conclusion of each interactant-utterance and the commencement of the echoborg's subsequent response. The average latency was 5.15 s ($SD = 3.04$ s).

Discussion. Our results indicate that under conditions of ambiguity wherein the source of an interlocutor's verbal agency is unclear, the interface substantially affects whether one attributes human agency to the words one's interlocutor produces. Participants who communicated with a chat bot via a text interface were significantly more likely to see their interlocutor as actually producing words generated by a chat bot compared to those who encountered the same chat bot but through a human shadower. The results from this study corroborate the notion that the cyranic illusion is robust in circumstances involving extreme source-shadower incongruity: people are biased toward perceiving an echoborg as an autonomous person.

Our findings suggest that it is relatively easy to get a chat bot to be perceived as an autonomous human if one is free to manipulate the contextual frame (i.e., the social psychological context of the interaction). An ostensibly simple suggestion from the

experimenter (i.e., that an interlocutor might be a human imitating a chat bot) can shift the entire contextual frame, fundamentally altering attributions of agency. Indeed, whenever it is claimed a certain bot has “passed the Turing Test” or some variant of Turing’s game, it usually has less to do with advances in conversational agent technology and more to do with shifting the contextual frame (e.g., when the chat bot Eugene Goostman—a bot that poses as a 13-year-old Ukrainian boy with limited English skills and general knowledge—was declared as having successfully fooled 33% of interrogators in a Turing Test in 2014; You, 2015). This, however, raises a fundamental question: within what contextual frame should participants encounter chat bots when we evaluate them? Arguably, the most important frame is the most common, namely, the everyday assumption that our interlocutors are human, just like us.

Study 3: Can covert echoborgs pass as human in the everyday contextual frame?

Aims. Study 3 examined people’s impressions following their conversing with an agent who, unbeknownst to them, produced solely the words of a chat bot. We aimed to gauge whether or not being shadowed by a human improved a chat bot’s ability to pass as an actual person within the everyday contextual frame (i.e., under the conditions of a generic social encounter wherein it is assumed an interlocutor is an ordinary human). The concept of “passing” within such a frame comes from the sociological and social psychological traditions that explore the mechanisms through which people manage identities in order to be accepted as a member of a particular group (Goffman, 1963; Renfrow, 2004; Khanna and Johnson, 2010). For example, the anthropomorphic androids in Dick’s (1968) novel *Do Androids Dream of Electric Sheep?* were able to pass as human so long as they concealed their true nature, took part in mundane human activities, and avoided the scrutiny of bounty hunters. The speech shadower in an echoborg is essentially a human mask placed over the peripherals one normally associates with computer systems. From a static third-person point of view, therefore, echoborgs appear to be autonomous human beings and nothing more, raising the question as to whether or not despite their communicative deficiencies people still sense that echoborgs are ordinary people. We predicted that research participants would not leave an interaction with a covert echoborg with the impression of having communicated with something non-human, whereas interacting with a covert chat bot through a text interface would leave participants with a strong impression of having encountered machine intelligence of some sort.

This study also investigated perceptual phenomena associated with the uncanny valley, namely how human-like, eerie, and familiar a covert echoborg interlocutor would seem to those with whom they communicated, and whether or not people would be comfortable in the presence of a covert echoborg. Mori's (1970) original hypothesis suggested that "subtle deviations from human appearance and behavior create an unnerving effect" (MacDorman and Ishiguro, 2006b, p. 299), and our goal was to gauge people's reaction to an interlocutor that was human in all respects but for the fact that a conversational agent determined the words they spoke.

Shadowers and subjects. A female graduate student (aged 23) was trained to perform as a speech shadower. Forty-one adult participants (26 female; mean age = 24.12, $SD = 7.59$) were randomly assigned to one of two conditions: Echoborg ($n = 20$) and Text Interface ($n = 21$).

Procedure. In addition to Cleverbot, two other chat bots were used in this study: Mitsuku (winner of the 2013 Loebner Prize) and Rose (winner of the 2014 Loebner Prize). In the Echoborg condition, Cleverbot and Rose were each assigned to speak with seven participants while Mitsuku spoke with six participants. In the Text Interface condition, Cleverbot, Rose, and Mitsuku each spoke with seven participants. During Cleverbot trials, the stock responses used in the prior two studies were employed.

The participant was instructed that the study concerned how strangers conversed when speaking for the first time, that it involved simply holding a 10-min conversation with another research participant, and that they were free to decide on topics for discussion so long as vulgarity was avoided. The researcher made no mention of chat bots or of anything related to artificial intelligence. Furthermore, the participant was given no indication that their interlocutor would behave non-autonomously or abnormally. The aim was to invoke the everyday contextual frame, in so far as that can be done within an experimental setting.

This study used the same minimal technological dependency apparatus and procedure as in Study 2. In the Text Interface condition the participant spoke aloud to their interlocutor while their interlocutor's responses were shown in text on a computer screen. In the Echoborg condition the participant encountered a human shadower face-to-face.

Measures and post-interaction interview. Following the interaction the participant

completed a brief questionnaire containing items asking them to indicate on a 10-point scale how human-like (1: very mechanical and computer-like; 10: very human-like), eerie (1: not at all eerie; 10: very eerie), and familiar (1: not at all familiar; 10: very familiar) their interlocutor seemed, as well as how comfortable they felt during the interaction (1: not at all comfortable; 10: very comfortable). Participants were also asked to briefly describe in writing the person they spoke with and what they thought they study was about.

When the questionnaire was completed, the researcher interviewed the participant to gain a sense of their impressions of the interaction and their interlocutor. The participant was asked to describe salient aspects of their interlocutor's personality. In order to ascertain whether the participant had picked up on the fact that they had communicated with a computer program, the researcher asked the participant whether they had suspicions regarding the nature of their interlocutor or about the study generally. Finally, the researcher revealed to the participant the full nature of the interaction and disclosed the purpose of the study.

Results. In the Text Interface condition, 14 of 21 participants (67%) mentioned during their post-interaction interview (prior to the researcher making any allusion to chat bots or anything computer-related) that they felt they had spoken to a computer program or robot. Two participants stated during debriefing that they suspected their interlocutor was a real person acting or using a script. Furthermore, seven participants (33%) explicitly stated in writing on their questionnaires that they believed the purpose of the study was to assess human-computer/human-robot interaction. Of the 14 participants who did not indicate that they thought the purpose of the study involved human-computer interaction, six said that they thought the study concerned how strangers communicated with one another (the stated purpose of the study supplied by the researcher prior to the interaction). Two participants believed the study concerned how people handle abnormal/unexpected situations. Six participants provided unique responses that did not fit into these categories.

Only 3 of 20 participants (15%) in the Echoborg condition stated during their post-interaction interview that they felt as though they had spoken to a computer or robot. Fifteen participants made it clear to the researcher during their interview that they suspected their interlocutor had been acting or giving scripted responses that did not align with their actual persona. Only two participants (10%) indicated in writing on their questionnaires that they believed the purpose of the study was to assess human-computer/human-robot interaction. Of the 18 participants

who did not indicate that they thought the study's purpose was to investigate human-computer interaction, only one stated that they thought the purpose of the study was to investigate communication between strangers. Seven participants believed the purpose of the study related to how people deal with abnormal/unexpected situations (e.g., "how people react when thrown out of their comfort zone" and "how people react to people who do not comply with social norms"). Four participants believed the study's purpose was to see how people communicated those who were shy/introverted. Three participants stated that they thought the study's purpose involved how people communicate with those who have a disability such as autism or speech impairment. Four participants provided other unique responses.

We performed a multivariate analysis of variance to see whether Interface (Echoborg vs. Text Interface) and Chat Bot (Cleverbot vs. Mitsuku vs. Rose) produced effects on participants' judgments concerning the four questionnaire items that pertained to how familiar, eerie, and human-like their interlocutor seemed as well as how comfortable they felt during the interaction. An initial omnibus test showed a significant effect of Interface, $\Lambda = 0.73, F(4,34) = 3.18, p < 0.05, \eta^2 = 0.27$, and a non-significant effect of Chat Bot, $\Lambda = 0.74, F(8,68) = 1.41, p = 0.21, \eta^2 = 0.14$. Univariate tests showed a significant effect of Interface on how comfortable participants felt during the interaction, $F(1,37) = 10.64, p < 0.01, \eta^2 = 0.22$, with participants in the Text Interface condition reporting higher levels of comfort ($M = 5.52, SD = 2.42$) compared to those in the Echoborg condition ($M = 3.44, SD = 2.04$). However, these univariate tests showed non-significant effects of Interface with respect to how familiar, $F(1,37) = 1.52, p = 0.23, \eta^2 = 0.04$, eerie, $F(1,37) = 0.08, p = 0.77, \eta^2 < 0.01$, and human-like, $F(1,37) = 0.24, p = 0.63, \eta^2 = 0.01$, interlocutors seemed. In the Text Interface condition, mean scores for familiarity, eeriness, and human-likeness were 3.81 ($SD = 1.89$), 6.19 ($SD = 2.14$), and 2.95 ($SD = 1.63$), respectively, compared to scores of 3.00 ($SD = 2.22$), 6.00 ($SD = 2.00$), and 2.70 ($SD = 1.78$), respectively, within the Echoborg condition.

Two Echoborg condition trials for each chat bot were selected at random and the audio latency was assessed. The average latencies for Cleverbot, Mitsuku, and Rose were 4.43s ($SD = 2.92s$), 5.95s ($SD = 3.98s$), and 3.96s ($SD = 3.94s$), respectively. As each trial made use of the same minimal technological dependency format of interaction, the differences between these latencies can be accounted for by the fact that the chat bots we used differ in terms of the speed at which they generate and return responses.

Discussion. In line with our hypothesis, a majority of participants in the Text Interface condition sensed they were communicating with a chat bot despite being led to believe they would be talking to another research participant while only a small minority of participants in the Echoborg condition came to the same conclusion. These results suggest that a chat bot stands a far greater chance of passing as a human in an everyday contextual frame when being shadowed by a human than when communicating via a text interface. The caveat to these findings, however, is that interactants do not tend to see a person shadowing for a chat bot as genuine. Rather, interactants see such people as deliberately behaving outside of their normal persona. This finding corroborates the general phenomenon observed in Study 2, that people are inclined to perceive an echoborg as somebody acting but nonetheless speaking self-authored words. We should note, however, that participants' awareness of being in a laboratory study may have contributed to their suspecting that the persona they encountered was not genuine. Future research may include observational field studies wherein interactants encounter a covert echoborg in real-world social contexts (e.g., a generic social gathering). It is plausible that in such scenarios interactants would be less inclined to form the belief that an echoborg was someone deliberately acting outside of their normal persona.

Although our experiment only considered two types of interfaces as opposed to a continuum of interfaces ranging from the very-human to the very-mechanical, our results contribute a novel finding to the discussion surrounding uncanny valley phenomena. We found evidence that people feel significantly less comfortable speaking to a chat bot through a human speech shadower than they do speaking to the same chat bot through a text interface. General discomfort seemed to derive from the social awkwardness that arose due to the chat bot's violations of conversational norms. The effect of these violations appears to have been magnified in the Echoborg condition. It is likely that participants in the Echoborg condition held higher expectations about the level of understanding and rapport that would be reached and sustained during the interactions on account of their speaking face-to-face with another human being, for the physical body of the other is laden with social cues that evoke such expectations (Kiesler, 2005). Komatsu and Yamada's (2011) "adaptation gap" hypothesis suggests that when expectations are not met during interactions with agents (e.g., when the implied social capacity of an agent exceeds that actually experienced by a user), people's subjective impressions are affected. Accordingly, participants in the Echoborg condition may have felt more uncomfortable compared to their counterparts in the Text Interface condition partly due to their having higher pre-interaction expectations about the quality of

interlocution they would experience. What requires further study is the investigation of conditions within which participants are told prior to interacting with either an echoborg or a text interface that their interlocutor will be producing the words of a chat bot. Adding two such conditions to Study 3's design would allow one to observe whether the body of the other produces effect on feelings of comfort independent of pre-interaction expectations.

General discussion

We have introduced and demonstrated a new research method, a special type of cyranoid we call an echoborg. Echoborgs make possible interactions with artificial conversational agents that have truly human interfaces. Though an abundance of research has demonstrated various means of embodying machine intelligence in human form, from onscreen embodied conversational agents (e.g., Cassell et al., 2000; Krämer et al., 2009) to 3D agents in immersive virtual environments (e.g., Selvarajah and Richards, 2005; Bailenson et al., 2008) to tangible machine-bodied androids (e.g., Ishiguro and Nishio, 2007; Spexard et al., 2007), the echoborg stands apart from these other methods in that it involves a real, tangible human as the interface.

Study 1 compared a standard text-based version of the Turing Test to an echoborg version and found that although a chat bot's ability to pass a Turing Test was not improved when being shadowed by a human, being shadowed did increase ratings of how human-like the chat bot seemed. This effect of embodiment on human-likeness was unique to chat bot interlocutors, as human interlocutors in these tests were not seen as more human-like when their words were spoken by a human shadower, suggesting that a demonstrated capacity for human-level dialog may override the effect of human embodiment on perceptions of human-likeness in Turing Test contexts. Study 2 showed that in an ambiguous situation wherein participants were told that an interlocutor was either articulating words generated by a chat bot or merely imitating one, participants in a text interface condition were more likely to conclude that they had encountered the words of an actual chat bot than those who encountered an echoborg. The contrast between these two conditions provides evidence for (a) the robustness of the cyranic illusion, and (b) the notion that people's causal attributions align with what is most salient and least ambiguous to them. Study 3 explored the notion of passing and the uncanny valley in an ordinary, everyday contextual frame (i.e., the experimental context attempted to simulate a generic, unscripted, first-time encounter between strangers). Participants engaged with a covert chat bot via either a text interface or

an echoborg. When interviewed following these interactions, most of the participants who engaged a text interface suspected they had encountered a chat bot, whereas only a few of the participants who engaged an echoborg held the same suspicion. This suggests that it is possible for a chat bot to pass as fully human given the requisite interface, namely an actual human body, and a suitable contextual frame. This study also found that people were less comfortable speaking to an echoborg than to a text interface.

Implications. *Android science.* Drawing from Nunamaker et al.'s (2011) distinction between virtual avatars and embodied conversational agents, in Figure 3.4 we visualize a simple two-dimensional matrix differentiating the basic tools available to android science, with one dimension indicating the source of verbal (and potentially non-verbal) agency and the other indicating interface-type. This matrix places the echoborg in relation to current mechanical devices utilized by android researchers (autonomous and tele-operated androids) as well as human beings as experimental subjects. By juxtaposing the field's tools in this manner, we can begin formally distinguishing the unique research questions that lend themselves to each. The fundamental question that each of these tools can be applied to concerns what happens when the human elements of an interlocutor are removed and replaced by artificial imitations. The unique questions that can be approached via the usage of echoborgs concern how real human bodies (not mere mechanical imitations) fundamentally alter people's perceptions of and interactions with machine intelligence.

In the echoborg paradigm, the communicative limitations of chat bots and other types of conversational agents are not treated as problematic barriers to fluid conversation. Rather, these limitations are directly operationalized; how the human body as an interface mediates the perception of these communicative limitations is what is of interest. We can thus differentiate the echoborg paradigm from the tele-operated android paradigm in the following manner. Tele-operated android research targets the social dynamics between humans and human-like machine interfaces. Given that conversational agents are relatively poor communicators, the tele-operated paradigm cedes speech-interpretation/generation responsibility to a human operator, whose experiences operating an android can also be the subject of inquiry. By contrast, the echoborg paradigm is interested in the social dynamics that emerge when the words artificial systems produce are refracted through actual human bodies during face-to-face interaction.

Figure 3.4: *Basic tools of android science.*

PHYSICAL INTERFACE: "BODY"		SOURCE OF VERBAL (AND/OR BEHAVIORAL) AGENCY: "MIND"	
		Computer Program	Human
Human-Like Android	Autonomous Android	Tele-Operated Android	
	Echoborg	Real Human / Cyranoid	

The affordance which grants the echoborg particular promise as a methodology is that it allows researchers the opportunity to study interactions under conditions wherein people believe they are speaking to an autonomously communicating person. The echoborg can interact covertly (i.e., without interactants expecting that they are communicating with a bot). Of course, chat bots and other conversational agents can be deployed covertly via traditional text interfaces—and many are (e.g., posing as real people in chat rooms, web forums, and social media websites in order to distribute marketing messages and collect user-data; Gianvecchio et al., 2011; Nowak, 2012). But as Study 3 shows, focused interaction with a covert chat bot via a text interface for a sustained period of time is very likely to result in the interactant sensing that they are not speaking to an actual person. Today's chat bots simply fail to sustain meaningful mixed-initiative dialog, and unless their words are vocalized by a tangible human body, their true nature is quickly exposed.

The Turing Test paradigm (and passing). Over half a century since its conception, the Turing Test paradigm remains a substantial area of interest in artificial intelligence and philosophy of mind. The usefulness of the Turing Test as a technological benchmark, its rules, and what it would mean for a machine to pass such a test (i.e., what, exactly, passing would be evidence of) are issues that have been hotly debated (e.g., Searle, 1980; Copeland, 2000; French, 2000; Harnad, 2000; Chomsky, 2008; Watt, 2008; Proudfoot, 2011). The non-philosophical literature on the Turing Test focuses largely on the technological aspects of candidate conversational agents (e.g., whether they occasionally make spelling mistakes) and the conditions that give rise to increased fooling (e.g., knowing vs. not knowing of the possible presence of a machine intelligence; Saygin and Cicekli, 2002; Gilbert and Forney, 2015). What remains to be explored in sufficient depth are the social psychological dynamics within standard and modified Turing Test scenarios: causal attributions, identity and power

relationships, questions asked and avoided, misunderstandings recognized and repaired, intersubjective achievement, and so on (e.g., Warwick and Shah, 2015). Our position is that the Turing Test is most useful when its orthodox interpretation is relaxed and it is applied not toward assessing the capacities of chat bots *per se*, but toward investigating aspects of human social nature. Indeed, the chat bot itself may be the least interesting element within a Turing Test scenario. A chat bot can be made to fool a human interrogator if the expectations of the interrogator are manipulated (e.g., through ambiguous framing). What is interesting is exploring the ways in which the chat bot's utterances interact with the interrogator's expectations, all within a particular contextual frame, so as to produce a social interaction that feels more or less comfortable or human.

In essence, the three studies we have presented are all modified Turing Tests in that they explore passing in one form or another (with Study 1 bearing the closest resemblance to Turing's original concept). What our studies show is how intimately connected passing is to the social psychological framing of an interaction, and how the interface one communicates with affects the meaning of the situation from the point-of-view of interactants. In our own view, the results from Study 3 are at the same time the most profound and the least surprising. Seventeen of 20 people spoke face-to-face with an echoborg in a small room for 10-min and failed to develop even the slightest suspicion that they were interacting with the words of an artificial agent of some kind. They may have seen their interlocutor as strange, introverted, or even acting, but it did not cross their minds that who (or what) they were dealing with was part computer program. This makes sense in light of how we experience mundane human interaction, and implies that, given certain generic social psychological preconditions, an interlocutor's capacity to produce sophisticated or even sensible syntax simply does not factor in to our categorizing them as a human being or as having a "mind." That is to say, rather than taking these results as indicating the sophistication of chat bots, we take these results as indicating the importance of both the body and social psychological framing in social interaction.

Future research applications. Creating human-like interfaces that totally override people's awareness that they are interacting with something artificial remains a distant holy grail (Vogeley and Bente, 2010). In the interim, however, we can use echoborgs to approximate the conditions of a world in which machines are capable of passing the nonverbal and motor requirements of a Total Turing Test. This opens the doors to a new frontier of human–robot

and human–agent interaction research.

Echoborgs can be used to further study uncanny valley phenomena. Most of the literature that has explored the uncanny valley has focused on motor behavior and physical resemblance as independent variables, as well as the effects different levels of participant engagement (passive vs. active) have on perceptions of agents (e.g., von der Pütten et al., 2011). Researchers have also, but to a lesser extent, looked at the role of phonetic quality in relation to the uncanny valley (e.g., Mitchell et al., 2011; Tinwell et al., 2011). Echoborgs enable us to study uncanny valley phenomena isolating dialogic capacity as an independent variable. Using echoborgs, we can see if an uncanny valley emerges when a spectrum of conversational agents ranging from the very poor (machine-like) to the very advanced (human-like) are communicated through a human speech shadower in unscripted face-to-face interactions.

Another possible avenue of research concerns the use echoborgs in comparative person perception studies. Experiments can be designed with conditions differentiated in terms of the interface through which participants communicate with a particular conversational agent (text interface, embodied conversational agent, echoborg, and so on). Researchers could then observe how the various interfaces shape aspects of the personality perceived by the participant, from minimal interfaces all the way up to a face-to-face human body.

A particularly enticing possibility for future research involves developing bots that simultaneously dictate words to a shadower while directing elements of the shadower’s motor behavior. In the echoborgs we have thus far constructed, the bot supplies the speech shadower with what to say while the shadower retains full control over their non-verbal functioning. We can imagine, however, developing a bot that delivered to the shadower’s left ear monitor words to speak while delivering basic behavioral commands (e.g., “smile,” “stand up,” “extend right hand for handshake”) to the shadower’s right ear monitor. This would grant the bot greater agency over the echoborg’s behavior.

The exciting opportunity opened up by echoborgs more generally is the opportunity to study human–computer interaction under the conditions of face-to-face human–human interaction. The problem for human–computer interaction research in general, and android science in particular, is that humans approach human–computer interaction differently from human–

human interaction (as our own research shows). Human–human interaction triggers a huge range of complex phenomena, from identity dynamics to social emotions to basic taken-for-granted assumptions to an incredibly subtle intersubjective orientation to the other (Gillespie and Cornish, 2014). The echoborg method enables us to test conversational agents within face-to-face interaction scenarios, simultaneously pushing AI into a new domain and also to probing the full complexity of the human–human inter-face.

Ethical considerations. In exploring social contexts involving a covert echoborg, mild deception is required in order to preserve the participant’s belief that they are encountering an autonomous person. Careful experimental design (e.g., choice of conversational agents and shadowers, duration of interaction, communicative setting, etc.) and thorough piloting of procedures is strongly recommended so as to render participant distress unlikely. Participants should be exhaustively debriefed to gauge whether or not adjustments need to be made to the research procedures in order to avoid potential negative experiences. As a guideline, the debrief procedure in Study 3 involved asking the participant if they had any concerns regarding the ethics of the study as well as if they would object to a close friend or relative taking part in the same study under the same conditions. All participants said no to both questions. We can anecdotally report that all of our participants enjoyed taking part in our research, with many expressing positivity toward the echoborg concept during debriefing and linking their experiences with what they had seen in popular science fiction films.

Limitations. Our studies were highly exploratory in nature. As such, various aspects of our investigations could have been more finely controlled. Though best attempts were made to standardize the body language of shadowers across all experimental trials, we did not make specific considerations for controlling certain behaviors (in particular, consistency of eye-contact). Moreover, the identity features of the shadowers (e.g., gender, ethnicity, age, and so on) may have produced unobserved effects on participants. We did not formally investigate such effects as they were not deemed to be of theoretical interest; however, we do acknowledge that questions regarding the relationship between the physical identity of the shadower and the social perception of the echoborg warrant future investigation. Sample sizes in our studies were relatively small due to practical constraints. Had our sample size for Study 3 been larger we might have been able to conduct a comprehensive comparison between the three chat bots used (Cleverbot, Rose, and Mitsuku). Also, we disclose that our choice of chat bots was based on prior familiarity with these programs.

We did not systematically analyze the effects audio latency may have had on participants' experiences. The delay between interactant-utterances and echoborg-responses in the studies that involved participants speaking aloud to an echoborg certainly degraded the mundane realism of interactions to some degree. Minimizing this latency is a major research priority as we continue to refine the echoborg methodology. At the moment we face a trade-off between speed and accuracy: the use of a speed-typing third party (the minimal technological dependency model) slows the pace at which the conversational agent receives the words spoken by the interactant, yet better guarantees that the agent will process an accurate representation of the interactant's words.

Conclusion. This article has demonstrated the possibility and potential of echoborgs: human-bodied entities whose words (and potentially motor actions) are partially or completely determined by a computer program. Researchers can use echoborgs to study how people interact face-to-face with machine intelligence under the assumption that it is human. This methodology opens up a new paradigm for human-computer interaction research as to date people have interacted with computers, even sophisticated agents and highly lifelike androids, as machines (i.e., as things categorically different from real humans). Pairing a conversational agent with a human being to create an echoborg fundamentally transforms how people perceive and emotionally experience an in person encounter with social technology. Perhaps the most exciting takeaway from this initial examination of echoborgs is that under certain social psychological conditions echoborgs pass as fully autonomous human beings.

CHAPTER 4 | Using echoborgs to assess intersubjective effort in human-agent dialog

Preface

Having explored the possibility of an echoborg extension of the cyranoid method in the previous chapter, we now turn toward using the echoborg to investigate a fundamental intersubjective human social behavior: the repair of misunderstandings in dialog. Whereas Chapter 3 is largely a proof of concept of the echoborg and a contribution to the field of android science, the current chapter employs the echoborg in a deep investigation certain factors that affect how much effort people exert toward attempting to achieve common ground with an artificial agent interlocutor. Via the cyranic illusion, the echoborg can create situations wherein an interactant *believes* they are speaking to an autonomously communicating person (being that they encounter another human body face-to-face). The current chapter operationalizes this affordance to demonstrate how the behaviors people express that are associated with generating shared understanding can be assessed in relation to manipulations regarding their knowledge of the true source of an interlocutor's words and the interface that they encounter (i.e., human body vs. machine).

The journal article that follows was authored by Corti and Gillespie. Corti designed and carried out the study, performed data analysis, determined third-turn coding criteria, outlined the article, and authored the main drafts, contributing roughly 75% of the content. Gillespie provided the technological devices, helped conduct the studies, provided key supervisory assistance, provided funding, provided Corti with editorial suggestions for the article, helped determine third-turn coding criteria, and contributed roughly 25% of the content. The article has been accepted for publication pending minor revisions in the journal *Computers in Human Behavior*. Chapter 4 includes these revisions.

Copies of participant instructions and other materials used to conduct the studies in this chapter can be found in Appendix F.

CHAPTER 4 | Using Echoborgs to Assess Intersubjective Effort in Human-agent Dialog

Article Title

Co-constructing intersubjectivity with artificial conversational agents: People are more likely to initiate repairs of misunderstandings with interlocutors represented as human.

Authors

Kevin Corti (primary)

Alex Gillespie

Abstract

This article explores whether people more frequently attempt to repair misunderstandings while speaking to an artificial conversational agent if it is represented as fully human. Interactants in dyadic conversations with an agent spoke to either a text screen interface (agent's responses shown on a screen) or a human body interface (agent's responses vocalized by a human speech shadower via the echoborg method) and were either informed or not informed prior to interlocution that their interlocutor's responses were agent-generated. Results show that people are less likely to initiate repairs with conversational agents when they communicate via a text screen interface and when they explicitly know their interlocutor's words to be agent-generated. That is to say, people demonstrate the most "intersubjective effort" toward establishing common ground when they engage the agent under the same social psychological conditions of face-to-face human-human interaction (i.e., when they encounter another human body and assume they are speaking to an autonomously-communicating person). The article's methodology presents a novel means of evaluating intersubjectivity in human-agent interaction against human-human benchmarks.

Keywords

Common ground, conversational repair, echoborg, human-agent interaction, intersubjectivity, psychological benchmarks

Introduction

"Intersubjectivity has [...] to be taken for granted in order to be achieved." –
Rommetveit (1974, p. 56)

Psychological research involving artificial agents designed to emulate human social capabilities (e.g., robots, androids, and conversational agents that interact using language and/or nonverbal behavior) has largely focused on whether people self-report the agent to be humanlike. Arguably, however, what is more important is whether the agent elicits humanlike patterns of interaction. Cassell and Tartaro (2007) claim that “the goal of human-agent interaction [...] should not be a believable agent; it should be a believable interaction between a human and agent in a given context” (p. 407). Accordingly, it has been proposed that the appropriate means of benchmarking an agent is to evaluate the extent to which the agent and a human interactant can together demonstrate a quality of intersubjectivity similar to that displayed in human-human interaction (Cassell & Tartaro, 2007; Schönbrodt & Asendorpf, 2011), herein referred to as demonstrating “benchmark intersubjectivity.” Intersubjectivity is a term that refers to the interactional relationship between perspectives within a dyad or larger group that becomes evident through each interactant’s behavioral orientation to the other (Gillespie & Cornish, 2010; Linell, 2009; Trevarthen & Aitken, 2001). Intersubjectivity is co-constructed within social interaction (Jacoby & Ochs, 1995), dynamically created through communicative processes within the interaction. When co-constructing intersubjectivity is used as a criterion for evaluating human-agent interaction (HAI), emphasis is placed not on isolated characteristics of either party (e.g., how humanlike the agent appears), but rather on the specific communicative processes through which the human-agent pair’s perspectives are coordinated.

A key intersubjective process demonstrated by humans involves the use of spoken language to build and sustain common ground (i.e., a shared understanding of the semantics and frames of reference particular to a given interaction) via a linguistic toolkit that enables the signaling, acknowledging, and repair of misunderstandings (Clark & Brennan, 1991; Schegloff, 1992). Merely possessing this toolkit, however, is insufficient for establishing common ground; this accomplishment requires active *facilitation* by each party to an interaction by-way-of regular and appropriate use of this toolkit (Altermann, 2007; Clark & Schaefer, 1989). When a person facilitates common ground at a level indicative of benchmark intersubjectivity, the person can be said to be exerting “benchmark intersubjective effort.” With respect to HAI, exerting benchmark intersubjective effort toward an agent is necessary otherwise the interactant will deprive the agent of the communicative support necessary to ascend into the complex intersubjective world of humans.

The current article tests the idea that absent the belief that they are engaging with an autonomously communicating person, human interactants will not exert benchmark intersubjective effort when in communication with an artificial agent, nor will they exert benchmark intersubjective effort when the agent communicates via a nonhuman interface (i.e., does not have a human body). This idea is explored via the “echoborg” method demonstrated by Corti and Gillespie (2015a). An echoborg is a hybrid entity composed of a human speech shadower who wears a concealed inner-ear audio receiver and vocalizes words they receive from a conversational agent. The technique enables social situations wherein people believe they are speaking to an autonomously communicating human (due to the fact that they engage with another human body face-to-face and in person) when in reality the words spoken by this human are entirely determined by an unseen conversational agent. This method can elicit an approximation of benchmark intersubjective effort from interactants in a baseline condition (i.e., human body interface + no explicit knowledge of an interlocutor’s words being agent-determined) that can be compared to the intersubjective effort demonstrated in conditions involving a nonhuman interface and/or explicit knowledge that an interlocutor’s words are agent-determined.

Intersubjectivity and intersubjective effort

Intersubjectivity has been conceptualized as entailing the interactions among (minimally) three levels of perspectives: (1) *direct*-perspectives (each party’s point-of-view), (2) *meta*-perspectives (what each party thinks the other party’s point-of-view is), and (3) *meta-meta*-perspectives (what each party thinks the other party thinks their point-of-view is) (Gillespie & Cornish, 2010; Ichheiser 1943; Laing, Phillipson & Lee, 1966). According to Gillespie and Cornish (2010), this framework can be used to understand social processes such as deception (i.e., the manipulation of meta-perspectives) as well as operationalize disagreement (i.e., misalignment between self’s direct-perspectives and other’s direct-perspectives) and misunderstanding (i.e., misalignment between self’s meta-perspective and other’s direct-perspective). This distinction between disagreement and misunderstanding is crucial: achieving common ground is *not* about parties agreeing with one another, but about parties forming accurate meta-perspectives in relation to the context of an interaction, and this is facilitated via empirically observable conversational processes that display and repair perspectives (see Clark & Brennan, 1991; Marková, 2003; Tirassa & Bosco, 2008).

Consider the following vignette, in which Aaron (from London) and Bryan (from New York)

have a conversation:

Aaron: How did you get to work today?

Bryan: I took the subway.

Aaron: You took the *subway*?

Bryan: Err, I mean I took the *underground*. I forgot that that's what you call it here in London.

Aaron: Got it.

Bryan formulates his initial response (“I took the subway”) on the assumption that Aaron’s direct-perspective with regard to the semantics of the utterance will match his direct-perspective (i.e., Bryan “designs” his utterance based on expectations he holds about Aaron; see Arundale, 2010; Gillespie & Cornish, 2014). Aaron then signals to Bryan that, in fact, he does not understand the semantics of Bryan’s initial response (“You took the *subway*?”), indicating that Aaron’s meta-perspective of the phrase “I took the subway” does not align with Bryan’s direct-perspective of the phrase. Bryan subsequently infers that Aaron is requesting an update to his meta-perspective and responds by clarifying the semantics of his initial response (“Err, I mean I took the underground. I forgot that that’s what you call it here in London”). As evidenced by Aaron’s final utterance (“Got it”), Bryan’s clarification sufficiently resolves the misunderstanding. Aaron now understands what Bryan meant by the phrase “I took the subway” as there is now alignment between Aaron’s meta-perspective and Bryan’s direct-perspective.

The intersubjective effort exerted by both Aaron and Bryan in pursuit of common ground is evidenced by the relationship between their various speech acts. Producing speech acts in support of establishing common ground is a process known as “grounding” (Clark & Brennan, 1991; Clark & Schaefer, 1987). At any fixed point in time prior to, during, and after a social interaction there exists a relationship between the various possible direct-, meta-, and meta-meta-perspectives held by each interactant. Behaviors arising from of intersubjective effort (e.g., grounding) cause these perspectives to act upon one another so as to make

evident to each interactant loci of agreement/disagreement and understanding/misunderstanding, and it is through such processes that the contents of perspectives are renegotiated and updated.

Analyzing intersubjective effort in dialog via repair activity. Conversation Analysis (CA) provides a basis for evaluating the quality of intersubjectivity in dialog (Gillespie & Cornish, 2010). CA arose out of the sociological tradition of “ethnomethodology” developed by Garfinkel (1967) and seeks to interpret language usage within the micro-context experienced by parties to an interaction (i.e., “talk-in-interaction”) rather than in its context-free, idealized form (Goodwin & Heritage, 1990; Hutchby & Wooffitt, 2008). Originators of the method identified fundamental organizational elements of talk-in-interaction, including how speakers allocate turns at talk as well as manage errors and misunderstandings (Sacks, Schegloff, & Jefferson, 1974; Schegloff, Jefferson, & Sacks, 1977), and CA has since proved useful in interactionist approaches to evaluating human-computer dialog (e.g., Brennan, 1991; Frohlich, Drew, & Monk, 1994; Raudaskoski, 1990; Zdenek, 2001). The current work focuses exclusively on the repair of misunderstandings, the mechanisms of which tie most directly to the operationalization of intersubjectivity and intersubjective effort described herein.

In the course of human dialog, interlocutors regularly produce utterances that contain errors or that are misunderstood. These utterances are often referred to as “trouble sources” (Schegloff, 1992). Repair activity is a type of grounding interactants deploy in order to mutually manage the presence of trouble sources and consists of the speaker of the trouble source (“self”) and the recipient of the trouble source (“other”) structuring their turns at speech so as to produce common ground. Successful repair sequences can take one of four general turn-taking forms (Zahn, 1984): (1) *self-initiated self-repair* involves the speaker of a trouble source both signaling and self-correcting a trouble source; (2) *other-initiated self-repair* involves the speaker of a trouble source self-correcting the trouble source following it being signaled by an interlocutor; (3) *self-initiated other-repair* involves an interlocutor correcting a trouble source following it being signaled by the speaker of the trouble source; (4) *other-initiated other-repair* involves an interlocutor both signaling and correcting the trouble source following its production by another speaker. These repair formats function as “the self-righting mechanism[s] for the organization of language use in social interaction” (Schegloff, Jefferson, & Sacks, 1977, p. 381), and according to Sidnell (2010), play “a vital

role in the maintenance of intersubjectivity" (p. 111).

Nearly all repair initiations occur within a "limited space around their self-declared ," while "virtually all *repairs* (i.e., solutions) occur within a very narrowly circumscribed space from their repair initiations" (Schegloff, 2000, p. 208, emphasis in original). There is a strong tendency for other-initiations of repair to occur in the turn following the utterance that contains the trouble source (the second position) and be immediately followed by a self-repair (Schegloff, 2000), creating a three-turn sequence known as "repair after next turn": (1) trouble source (self) → (2) repair initiation (other) → (3) repair outcome (self). As the third position provides the speaker of a trouble source an opportunity to resolve a misunderstanding in the brief window of space opened by the repair initiation, repair after next turn has been described as "the last structurally provided defense of intersubjectivity in conversation" (Schegloff, 1992, p. 1295).

In the terminology of Laing, Phillipson and Lee (1966), three turns are the minimal unit required to establish mutual understanding: the first turn presents a direct-perspective, the second turn questions that direct-perspective (i.e., it is a meta-perspective), and the repair, if successful, corrects or updates the meta-perspective. Repair after next turn thus coordinates perspectives, providing an elemental three-turn stitch in the co-created fabric of intersubjectivity.

Analysis of other-initiated self-repair can also be linked to intersubjectivity by considering how its mechanics involve bilateral joint attention, a prerequisite of complex intersubjectivity. When engaged in joint activities involving shared intentionality (i.e., the ability to understand a joint activity not merely from multiple subjective points-of-view, but also from a "bird's eye" point-of view from where the perspectives of self and other are seen as integrated; Tomasello, Carpenter, Call, Behne, & Moll, 2005), humans can through a repertoire of behaviors (e.g., speech acts) direct the attention of other humans to aspects of their shared environment relevant to shared goals (e.g., the goal of establishing common ground). Kaplan and Hafner (2006) outline four skills that an actor (biological or otherwise) must possess in order to accomplish bilateral joint attention: (1) *attention detection* (i.e., the ability to track the attention of others), (2) *attention manipulation* (i.e., the ability to manipulate and influence the attention of other actors through verbal and/or nonverbal gestures), (3) *social coordination* (i.e., the ability to engage in coordinated interaction with

others via techniques such as turn-taking and role switching), and (4) *intentional understanding* (i.e., the ability to understand the intentions of others and interpret and predict others' behaviors as they relate to goals).

Consider once again the following vignette (“TS” indicates a trouble source, “RI” indicates an other-initiation of repair, and “R” indicates a repair):

Aaron: How did you get to work today?

Bryan: TS → I took the subway.

Aaron: RI → You took the *subway*?

Bryan: R → Err, I mean I took the *underground*. I
forgot that that's what you call it here in
the U.K.

Aaron: Got it.

At work in this passage are each of the four requisite skills for bilateral joint attention outlined by Kaplan et al. (2006), thus the complexity of intersubjectivity between Aaron and Bryan, and the intersubjective effort exerted by both, can be observed. Bryan's first-position response (“I took the subway”) is misunderstood by Aaron. Aaron's misunderstanding is signaled in the next turn in the form of a repair initiation (“You took the *subway*?”) that functions as an attempt to focus Bryan's attention on the previous utterance wherein lies the trouble source (*attention manipulation*). As a direct consequence of the repair initiation, Bryan becomes aware of the fact that Aaron's attention is turned backward toward a trouble source located in Bryan's first position utterance (*attention detection*). Bryan infers that Aaron's intention in uttering the repair initiation is to elicit a third-position repair (*intentional understanding*), thus Bryan clarifies the trouble source in his next turn. The entire repair sequence occurs within a formal structure of turn-taking supported by both interlocutors (*social coordination*).

Kaplan et al.'s (2006) four requisite skills can be segmented into behavioral and non-behavioral varieties. Attention manipulation involves overtly producing a behavior intended

to influence the perspective of an interactant (e.g., uttering an other-initiation), while social coordination encompasses synchronizing one's behavior in accordance with that of an interlocutor in a manner conducive for the communication of perspectives (e.g., turn-taking). Attention detection and intentional understanding, meanwhile, are principally cognitive skills that do not necessarily manifest in the form of observable motor or linguistic behaviors (one can understand the intentions of another without producing an associated behavior). Insofar as intersubjective effort is operationalized as a behavioral indicator of a commitment to shared understanding, evidence for it in dialog can be found in observable actions such as other-initiations of repair. Failing to manipulate the attention of an interlocutor so as to alert them to the presence of a misunderstanding *when one otherwise could* constitutes a lack of intersubjective effort. For instance, had Aaron for whatever reason *not* uttered a repair initiation despite misunderstanding Bryan's use of the word "subway," then Bryan would have failed to recognize that his direct-perspective and Aaron's meta-perspective of the word "subway" were incongruent and the two interlocutors would thereby have failed to establish common ground.

Intersubjectivity in human-agent dialog: The role of interfaces and agency framing. Why might an interactant fail to exert benchmark intersubjective effort when in communication with an agent when they otherwise could? Answering this question requires considering how the agent is represented in the mind of the interactant. Specifically, it requires considering the factors that influence how the interactant generates meta-perspectives of the agent's direct-perspectives and how these perspectives are interacted with (if at all). This article examines two such factors: (1) the nature of the agent's interface (i.e., its embodied means of participating in social communication), and (2) the framing of the agent's communicative agency (namely, whether or not the interactant holds the belief that they are talking to an agent vs. another human being).

First, consider the role of interfaces in fostering intersubjectivity. The sense that an interlocutor possesses attention that can be manipulated so as to jointly manage misunderstandings provides to an interactant the impetus for intersubjective effort, and attributing attention to a potential interlocutor involves the supposition that said interlocutor has a subjective perspective of a shared social world (see Graziano, 2013). Detection of the subjective perspectives possibly held by another interlocutor involves inferring information signaled via the interlocutor's interface (e.g., its physical body), therefore the properties of an

interlocutor's means of interfacing influence how an interactant perceives and orients to the interlocutor's perspectives (be they real or imagined) and thereby influence the interlocutor's attention.

That an interface can exert such a powerful influence over intersubjectivity has long been of interest to psychologists and philosophers concerned with the embodied nature of perspective-taking (e.g., on this topic, the phenomenologist Husserl invoked the concept of "analogical apperception" – reflexively apperceiving other people's subjectivity based on their appearing to be similarly embodied and thereby becoming an "Other"; Husserl, 1931; also see De Preester, 2008; Hemberg, 2006). More recently, this connection between interfaces and the intersubjective relationship between two or more parties has been triangulated upon by numerous empirical research streams connected to social robotics and HAI. For example, in a neuroimaging study that involved humans interacting with a spectrum of entities ranging from extremely non-humanlike computers to humanlike androids to actual humans, Krach et. al (2008) demonstrated that "the tendency to build a model of another's mind linearly increases with its perceived human-likeness" (p. 1). Riek, Rabinowitch, Chakrabarti and Robinson (2009), meanwhile, found that people self-report greater empathy for robots perceived to be humanlike than for non-humanlike robots. Furthermore, Saygin and Stadler (2012) showed that people are more accurate at processing and predicting the motor behavior of humanlike vs. non-humanlike agents, suggesting that the degree to which the motor activity of an agent "resonates" with a human observer corresponds with how humanlike the agent is perceived to be. These findings suggest that as an agent's means of interfacing becomes more humanlike, the degree to which interactants consciously and unconsciously form a model of an agent's perspectives and attention increases (this is often referred to as "mentalizing," or demonstrating "theory of mind"). This also implies that the more an interactant's awareness of an agent's perspectives is reduced as a result of the agent's particular means of interfacing, so will be the interactant's impetus for exerting benchmark intersubjective effort toward the agent.

The notion that artificial agents with humanlike means of interfacing provide for more intersubjectively rich interactions has inspired the development of both embodied conversational interface agents (sometimes referred to simply as embodied agents, or intelligent virtual agents) and androids. Embodied agents are conversational agents that have been combined with anthropomorphic onscreen or immersive virtual interfaces. Many can

respond to both verbal and non-verbal input, generate verbal and non-verbal output, engage in repairs of misunderstanding, and communicate about the communication they are engaging in (Bailenson & Yee, 2005; Cassell, 2000). Androids, meanwhile, are physical machine imitations of humans. The field of android science has used such machines to better understand principles of human psychology being that the similarities in morphology between androids and humans allow researchers to investigate whether people respond in an alike manner when interacting with human and humanlike stimuli (Ishiguro & Nishio, 2007; MacDorman & Ishiguro, 2006). Android science has shown that while humans do demand more sociality from actual humans than from androids, people expect more sociality from androids than from mechanical looking robots and lesser-looking agents (MacDorman, 2006). The echoborg was introduced to the field of android science by Corti and Gillespie (2015a) in order to leapfrog current bottlenecks concerning the imperfect appearance and motor behaviors of contemporary androids as an echoborg approximates an android that can “pass” as human in terms of physical appearance and motor behavior.

An interactant’s mental formulation of the potential perspectives held by an interlocutor is not solely a function of the interlocutor’s means of interfacing. In fact, the meta-perspectives of an interlocutor’s direct-perspectives held by an interactant can be manipulated independently simply by altering the interactant’s beliefs about the interlocutor. Indeed, many experiments that assess the degree to which people engage with the real or imagined perspectives of other entities involve varying the ways in which an entity’s communicative agency is framed so as to induce changes in intersubjectivity. In HAI research this often entails either priming research participants to believe that they are engaging a fully-autonomous agent when in reality the agent is human-controlled, an approach commonly referred to as the “Wizard of Oz” technique (Dahlbäck, Jönsson, & Ahrenberg, 1993), or priming them to believe they are engaging a real person when they are in reality interacting with an agent. Studies that have adopted this approach have shown that people mentalize less about an entity with whom they interact when they believe the entity to be controlled by an artificial agent rather than an actual person (Chaminade et. al, 2012; Gallagher, Jack, Roepstorff, & Frith, 2002; Kircher et. al, 2009; also see Branigan, Pickering, Pearson, McLean, & Brown, 2011). Furthermore, Kennedy, Wilkes, Elder, and Murray (1988) found that in the context of text-based human-agent dialog, the primed belief that an interlocutor was actually a real person led to an increase in interactants’ use of anaphors (words that point back to earlier parts of a conversation), implying that people less often attempt to direct an

interlocutor's attention backward toward prior utterances when they believe their interlocutor to be a nonhuman agent. This suggests that intersubjective effort can potentially be impacted by the mere belief that one is or isn't speaking with another human being.

While with traditional HAI methods researchers can prime the belief that an agent is really a human, this approaches is only possible when used in conjunction with a nonhuman interface (i.e., a researcher cannot convince a research participant that a robot is actually an autonomous human being, they can only prime the belief that the robot is controlled by a real person, be that true or false in reality). Although embodied agents and androids mimic human likeness in a manner that augments the complexity of intersubjectivity expected by interactants, these interfaces are not fully human and, therefore, do not evoke the full spectrum of intersubjective expectations that color true human-human interaction (MacDorman, 2006). HAI research, therefore, has never to-date investigated HAI within a fully human-human social psychological frame wherein *both* the means of interfacing is fully human and the implied agency of the interlocutor is fully human. Since the echoborg method of HAI can achieve this, it presents a way to investigate the intersubjective processes that occur between an interactant and an artificial agent when the interactant both believes they are speaking to an autonomous human being and encounters a truly human interface (Corti & Gillespie, 2015a).

Experimental study

Overview. The following study assessed instances of other-initiated self-repair in dyadic conversations between research participants (interactants) and the artificial conversational agent Cleverbot, a text-based chat bot developed by Carpenter (2015). The study explored whether interactants' conversational repair behavior changes depending on whether an agent-interlocutor communicates through an actual human body (as opposed to a text screen interface) and whether the interactant explicitly knows their interlocutor to be communicating the words of an agent. A 2×2 experimental design was utilized with the factors *Screen* (1: Cleverbot communicated with the interactant via text on a computer screen; 0: Cleverbot communicated with the interactant via a human speech shadower - an echoborg) and *Aware* (1: the interactant was informed before the interaction that their interlocutor's words were those of a chat bot; 0: the interactant was *not* informed before the interaction that their interlocutor's words were those of a chat bot). The study was approved by an ethics review panel at a major British university and conducted in a behavioral research

laboratory.

Hypotheses. The study tested four hypotheses predicting main effects of the factors *Screen* and *Aware* on two separate dependent measures related to interactants' intersubjective effort: (1) interactant other-initiations (other-initiations produced by the interactant following a Cleverbot utterance) and (2) interactant self-repair attempts following Cleverbot other-initiation. These hypotheses were developed based on the argument that interactants' intersubjective effort would be greatest in the "covert echoborg" baseline condition that featured Cleverbot interacting through a human speech shadower and interactants not being informed prior to the conversation commencing that their interlocutor's words would be determined by an artificial agent. Regarding interactant other-initiations, it was predicted that interactants would be less likely to produce an other-initiation following a Cleverbot utterance when speaking to via a text screen interface (Hypothesis 1) and when explicitly aware that their interlocutor's words were determined by a conversational agent (Hypothesis 2). Likewise, it was predicted that interactants would be less likely to produce a self-repair following a Cleverbot other-initiation when speaking to Cleverbot via a text screen interface (Hypothesis 3) and when explicitly aware that their interlocutor's words were determined by a conversational agent (Hypothesis 4).

Participants (interactants) and shadower. In total, 108 adults (69 female; mean age = 25.87, $SD = 8.35$) participated in the study and were randomly assigned to experimental conditions. These interactants were recruited online via a university research participant recruitment portal and consisted of London-based university students, university employees, and adults unaffiliated with the university. A female graduate student (aged 30) functioned as the speech shadower in the two conditions that involved interactants engaging a human interface.

Procedure and apparatus. Following informed consent, the interactant was taken to an interaction room where they were instructed by the researcher as to how the study would proceed. The interactant sat in a chair at one end of the room and was told that the study involved speaking to an interlocutor for 10-minutes. The interactant was informed that they could decide for themselves topics to discuss with their conversation partner so long as nothing was vulgar. The non-scripted nature of the interaction was emphasized in order to allay any suspicions that the interlocutor would be speaking rehearsed responses. The

procedures for the separate experimental conditions were as follows:

“Covert echoborg” scenario: (*Aware* = 0, *Screen* = 0). The interactant was informed that the interlocutor (the female speech shadower) would enter the interaction room and sit in a chair facing the interactant shortly after the researcher exited the room, and that the interlocutor would initiate the conversation. Although the interlocutor would be shadowing words generated by Cleverbot in response to things the interactant said, this fact was not made known to the interactant at any point prior to or during the interaction, and the researcher made no allusion to conversational agents or chat bots prior to the interaction commencing.

“Overt echoborg” scenario: (*Aware* = 1, *Screen* = 0). As with the covert echoborg scenario, the interactant was informed that their interlocutor would enter the interaction room and initiate a conversation shortly after the researcher exited. Prior to exiting, however, the researcher informed the interactant that this interlocutor would be wearing an inner-ear device and would speak aloud words they received from a chat bot computer program located in an adjacent room. It was made clear to the interactant that the speech shadower would not speak any of their own thoughts during the interaction and that only the chat bot would respond to words the interactant spoke.

“Covert text bot” scenario: (*Aware* = 0, *Screen* = 1). The interactant sat facing a computer monitor on which a blank instant messaging client (Pidgin) dialog box was displayed. The interactant was informed that though they would speak aloud to their interlocutor, their interlocutor would respond via text that would appear on the monitor. As with the covert echoborg scenario (above), the interactant was not informed that their interlocutor’s words were determined by a chat bot and no

allusion to conversational agents or chat bots was made by the researcher. The researcher informed the interactant that shortly after they exited the room their interlocutor would initiate the conversation.

“Overt text bot” scenario: (*Aware* = 1, *Screen* = 1). As with the covert text bot scenario, the interactant sat facing a computer monitor on which a dialog box appeared and was instructed that though they would speak aloud to their interlocutor, their interlocutor would respond via text readable on the monitor. As with the overt echoborg scenario (above), the interactant was told that their interlocutor’s words would be those of a chat bot and that the interlocutor would initiate the conversation shortly after the researcher left the room.

The experimental apparatus was identical to that described by Corti and Gillespie (2015a) in their demonstration of minimal technological dependency interactant ↔ chat bot audio relay (for a video demonstration, see Corti, 2015). From a room adjacent to the interaction room, the researcher listened to the interactant’s speech via a “bug” microphone placed near the interactant and speed typed the interactant’s words into the Cleverbot program. In conditions involving the interactant engaging an echoborg (*Screen* = 0), the researcher spoke Cleverbot’s responses into a microphone which relayed to a discreet inner-ear monitor worn by the shadower, whereas in conditions involving the interactant engaging a computer interface (*Screen* = 1), the researcher relayed Cleverbot’s responses to the interactant’s computer monitor via the Pidgin instant messaging client. In their use of a minimal technological dependency interactant ↔ Cleverbot audio relay scenario, Corti and Gillespie report an average latency (the time between the conclusion of an interactant utterance the production of a Cleverbot response) of 5.15 seconds. In all conditions, the researcher relayed the phrase “hi there” to the shadower/screen to initiate the conversation. In order to establish identity consistency between trials, several stock responses were used in lieu of Cleverbot’s actual response for certain interactant utterances. When interactants inquired as to where their interlocutor was from, the stock response “I’m from London” was provided. If the interactant inquired as to their interlocutor’s occupation, the stock response “I’m a student here,” and if the interactant asked what their interlocutor studied, the stock response “psychology” was

used. Finally, if the interactant asked their interlocutor what their name was, the stock response “Kim” was provided. The shadower was instructed to maintain a consistent nonverbal demeanor across trials that reflected the spirit of the words generated by Cleverbot and to maintain eye-contact with the interactant during vocal delivery.

Measures: Coding and quantifying intersubjective effort. Following the conclusion of all experimental trials, transcripts of the interactions were prepared based on Cleverbot’s textual input/output logs. Each was assigned a random identification number so that they could be coded without the coder knowing the experimental condition to which a transcript belonged.

Testing each hypothesis required quantifying instances of other-initiated self-repair activity evident in each experimental trial. Researchers who use CA rarely quantify the phenomena they study, however Schegloff (1993) does offer guidance on how one might proceed with such an undertaking. A key to quantifying within the spirit of CA, according to Schegloff (1993), is properly identifying “environments of relevant possible occurrence” (p. 103), these being the locations within dialog where certain speech acts are likely to be located. In the case of repair initiations, such environments are clearly defined given that any utterance can act as a potential trouble source (ten Have, 1999); other-initiations of repair, thus, “can in principle occur *after any turn at talk*” (Schegloff, 1993, p. 115, emphasis in original). Environments of relevant possible occurrence are likewise well-defined for instances of third-position self-repair outcomes as they occur in the turn following other-initiation. It is important to note that while any turn at talk can potentially act as a trouble source, trouble sources themselves cannot be identified in isolation (i.e., they cannot be identified unless they are followed by a repair initiation). Trouble sources, thus, are “launched” from the second-position (Schegloff, 2007).

Criteria articulated by Schegloff et al. (1977) and Sidnell (2010) were used to establish what instances of talk counted as an other-initiation of repair. Other-initiation can involve the use of question words (e.g., *Huh?* *What?* *Who?* *Where?* *When?*), partial repeats of the trouble source (e.g., *The subway?*), and full repeats of the trouble source (e.g., *You took the subway?*) alone or in combination with one another, as well as demonstrations of possible understanding (e.g., *You took the subway... the walkway beneath the street?*). Other-initiations can be and often are explicit in declaring the presence of a misunderstanding (e.g., *I don’t understand; I don’t get what you just said; I’m not following you*). Repair initiations

that treat the whole prior turn as a trouble source rather than reference a particular element within the prior turn are known as “open” class repair initiators (Drew, 1997) and often take the form of single-word utterances (e.g., *Pardon?*). In some instances, other-initiations are triggered by failure to hear or mishearing words spoken by an interlocutor (Schegloff et al. 1977; Zahn, 1984) and involve a request that the first-position speaker repeat a trouble source (e.g., *I'm sorry I didn't hear what you just said*). However, since half of the experimental conditions involved participants reading text rather than being spoken to audibly, instances of other-initiation that could be linked to problems of hearing were excluded from analysis.

Third-position interactant utterances (those following a Cleverbot other-initiation) were classified as either legitimate attempts at self-repair (i.e., utterances that acknowledged and attempted to clarify a trouble source) or as non-repairs (i.e., utterances that did not attempt to clarify a trouble source) on the basis of criteria gleaned from Schegloff et al. (1977) as well as Schegloff (1997). Repairs are usually “successful and quick” (Schegloff et al., 1977, p. 364). Successful third-position self-repairs often involve the speaker of a trouble source repairing the trouble source via rephrasing or elaboration. Generally, a logical relationship between the repair outcome and the trouble source is overt in instances of legitimate attempts at repair (e.g., *Err, I mean I took the underground*), while this relationship is often absent or ambiguous in instances of non-repair. Non-repair can take the form of overt repair abandonment (e.g., *Just forget it*) or the production of a non sequitur. Oftentimes a non-repair can be identified where the third-position utterance leads to a subsequent other-initiated repair, creating a connected chain (or “cascade”) of unresolved repair attempts.

On the basis of these criteria, the following classification codes were assigned to each turn-at-talk for each transcript: *Repair Initiation* (other-initiation), *Repair* (legitimate attempt at self-repair following other-initiation), *Non-Repair* (unsuccessful self-repair following other-initiation), and *Null* (turn-at-talk did not meet criteria for any other code).

Results

A second coder coded a random subset of the transcripts in order to establish interrater reliability (4 transcripts from each condition). High consistency was found among raters, Cohen's Kappa = 0.81, $p < 0.001$, 95% CI = [0.76, 0.86].

Interactant other-initiation behavior. A multilevel logistic regression model was used to

test Hypothesis 1 and Hypothesis 2 (level 1: experimental dyad; level 2: experimental condition), with each observation being a turn-at-talk taken by Cleverbot. The dependent measure was a binary variable that took the value of 1 if the turn-at-talk was followed by an interactant other-initiation, fixed factors were *Screen* and *Aware*, and random intercepts were conditioned on each experimental dyad (i.e., each unique trial). This model showed a significant main effect of *Screen*, $b = -0.38$, $SE = 0.18$, $p < 0.05$, odds ratio (OR) = 0.68, 95% CI $OR = [0.48, 0.97]$, supporting Hypothesis 1: engaging a text interface vs. a human body interface resulted in a 32% reduction in the odds that an interactant would respond to a turn-at-talk taken by Cleverbot with an other-initiation, all else being equal. The model also showed a significant main effect of *Aware*, $b = -0.52$, $SE = 0.19$, $p < 0.01$, $OR = 0.59$, 95% CI $OR = [0.41, 0.86]$, supporting Hypothesis 2: explicitly knowing that their interlocutor's words were determined by an agent vs. not explicitly knowing resulted in a 41% reduction in the odds that an interactant would respond to a turn-at-talk taken by Cleverbot with an other-initiation, all else being equal. No significant interaction was found between *Screen* and *Aware*, $b = 0.24$, $SE = 0.28$, $p = 0.39$. The model included 3,612 observations nested within 108 dyad groups, AIC = 3083, BIC = 3114, random effect variance = 0.24.

Interactant self-repair behavior. A logistic regression model was used to test Hypothesis 3 and Hypothesis 4, with observations being each other-initiation produced by Cleverbot. The dependent measure was a binary variable that took the value of 1 if the other-initiation was followed by the interactant legitimately attempting to repair the trouble source. This model showed neither a significant main effect for *Screen*, $b = -0.54$, $SE = 0.54$, $p = 0.32$ (disconfirming Hypothesis 3), nor *Aware*, $b = -0.42$, $SE = 0.60$, $p = 0.48$ (disconfirming Hypothesis 4), nor was there a significant interaction between these factors, $b = 0.70$, $SE = 0.87$, $p = 0.42$. The specification of random effects conditioned on each experimental dyad did not improve model fit. Though both hypotheses were disconfirmed on the basis of the model showing no statistically significant effects of the fixed factors, legitimate attempts at self-repair following Cleverbot other-initiations were, on average, most common in the baseline condition (covert echoborg). The model included 210 total observations. 18 dyad groups were excluded on the basis of their containing no Cleverbot-uttered other-initiations.

A closer look at the co-construction of intersubjectivity

Bot self-repairs (or the lack thereof) vs. Interactant self-repairs. On average, Cleverbot successfully repaired only 22.27% ($SD = 27.14\%$) of self-uttered trouble sources following

other-initiation. Below are two examples of third-position non-repair by Cleverbot (TS, RI and NR indicate trouble source, repair initiation and non-repair respectively):

Excerpt 1

(Covert echoborg scenario, trial 15):

Interactant: It doesn't matter what age you are.

Cleverbot: TS → Yes it does my life makes sense.

Interactant: RI → To age?

Cleverbot: NR → Do you have a father?

Interactant: Ya, I do. Do you have siblings?

Excerpt 2

(Overt text bot scenario, trial 17):

Interactant: Do you have any siblings?

Cleverbot: Yes.

Interactant: How many?

Cleverbot: TS → You left-handed, under-handed mess
of wires and batteries.

Interactant: RI → What?

Cleverbot: NR → Never mind.

Interactant: Can you ask me a question?

Excerpt 1 is an example a common type of Cleverbot non-repair observed in the dataset, that of Cleverbot producing a non sequitur following other-initiation rather than address the

trouble source. The non-repair in Excerpt 2, on the other hand, is a clear example of self-repair abandonment, as Cleverbot simply utters “never mind” rather than attempt a repair of the trouble source flagged by the interactant.

The legitimate attempts at self-repair Cleverbot did produce often took the form of yes/no clarification, this form of repair requiring less explicit elaboration of and engagement with the trouble source. Two examples of this type of repair are shown below (“R” indicates a self-repair of other-initiation):

Excerpt 3

(Covert text bot scenario, trial 16):

Cleverbot: Who are you married to?

Interactant: My wife.

Cleverbot: TS → Who is your wife?

Interactant: RI → Do you want her name?

Cleverbot: R → Yes.

Interactant: Elanor.

Excerpt 4

(Covert echoborg scenario, trial 29):

Interactant: Well, I do have exams coming up.

Cleverbot: TS → Really? How do they work?

Interactant: RI → Exams?

Cleverbot: R → Yes, they are very stupid.

Interactant: You go into a room, you do whatever

the paper says.

While Cleverbot only managed to produce legitimate attempts at self-repair in response to roughly a quarter of the other-initiations produced by interactants, interactants themselves managed a rate of 84.36% ($SD = 30.65\%$). Below are two examples of legitimate self-repair attempts made by interactants:

Excerpt 5

(Overt text bot scenario, trial 13):

Cleverbot: No, you're talking about Hamlet script.

Interactant: TS → I'm not familiar with Hamlet actually.
I'm talking about Nirvana.

Cleverbot: RI → I don't know anyone named Nonsense,
so I don't know what you're talking
about.

Interactant: R → Nirvana.

Cleverbot: I love Nirvana!

Excerpt 6

(Overt echoborg scenario, trial 1):

Cleverbot: Knock knock knockin' on heaven's
door!

Interactant: TS → You like that music?

Cleverbot: RI → Like what?

Interactant: R → Do you like the music knockin on

heaven's door?

Cleverbot: Yes.

As with Excerpt 5 and Excerpt 6, legitimate interactant self-repairs attempts often came in the form of elaborations or clarifications of the trouble source. Interactant non-repairs, meanwhile, typically involved the production of a non sequitur that initiated a new conversational trajectory, as shown in Excerpt 7 and Excerpt 8:

Excerpt 7

(Overt echoborg scenario, trial 2):

Interactant: So what's your family life like?

Cleverbot: I don't have one.

Interactant: TS → That might explain quite a bit then.

Cleverbot: RI → Might?

Interactant: NR → So are you running away from
something by being a student in
London?

Excerpt 8

(Covert text bot scenario, trial 1):

Cleverbot: Hello.

Interactant: TS → Good morning.

Cleverbot: RI → Is it morning where you are?

Interactant: NR → How are you?

Interactant repair initiation inhibition. As the evidence in support of Hypothesis 1 and Hypothesis 2 showed, interactants in the covert echoborg scenario *worked harder* to achieve common ground as they more frequently attempted to elicit repairs from their interlocutor in relation to interactants in other experimental conditions. This suggests that interactants in non-baseline conditions more frequently “passed” on opportunities to initiate repairs where a repair might have been warranted. Consider the following excerpts (“NS” indicates a non sequitur; “P” indicates a pass):

Excerpt 9

(Covert text bot scenario, trial 10):

Cleverbot: I think you’re interesting too.

Interactant: Thank you.

Cleverbot: You’re welcome.

Interactant: What do you think of the recent London tube strike?

Cleverbot: NS → Just a celestial being with the abilities to use divine intervention but does not use it for the danger or oblivion.

Interactant: P → Fair enough.

Excerpt 10

(Overt echoborg scenario, trial 20):

Cleverbot: What book should I read?

Interactant: Well, I don’t know. I don’t know what kind of literature you like.

- Cleverbot:** Why don't you know?
- Interactant:** Well, it's just because I don't know you.
- Cleverbot:** NS → I don't speak German, but my grandfather is German.
- Interactant:** P → Okay cool. Well then you... would you like to read a book with me?

In the excerpts shown above, the interactants choose not to initiate repairs following clear non sequiturs uttered by Cleverbot. In Excerpt 9 the interactant appears indifferent to Cleverbot's non sequitur and responds with an apathetic statement ("Fair enough"). A somewhat different approach is taken by the interactant in Excerpt 10; rather than initiate a repair of Cleverbot's non sequitur, the interactant attempts to steer the conversational trajectory back toward the prior topic that the interlocutors had been discussing.

Handling non-repair. Interactants dealt with non-repair by producing one of three general response-types: (1) *successive other-initiations*, (2) *continuations* and (3) *departures*.

Successive other-initiations were instances where an interactant requested that Cleverbot repair either the non-repair located in the previous turn (as in Excerpt 11, below) or the original trouble source (as in Excerpt 12, below), creating a connected series of adjacency pairs (a "repair initiation cascade"):

Excerpt 11

(Covert echoborg scenario, trial 24):

Cleverbot: TS → Because they are birds.

Interactant: RI → I'm sorry?

Cleverbot: TS → Don't be.

Interactant: RI → Excuse me? Don't be what?

Cleverbot: TS → You may be excused.

Interactant: RI → From this?

Cleverbot: TS → From this moment what?

Interactant: RI → What do you mean by that?

Excerpt 12

(Covert text bot scenario, trial 17):

Interactant: What's your favorite game?

Cleverbot: TS → He is a legend.

Interactant: RI → He is a legend? Is it a game?

Cleverbot: TS → Yes you are.

Interactant: RI → I'm a game?

Cleverbot: NR → I'm a cat.

Continuations, meanwhile, encompassed engaging with the content of a non-repair without requesting that it or the original trouble source actually be repaired (i.e., the non-repair is not treated as problematic). This activity is analogous to that which has been described by CA researchers who investigate interactions involving one or more persons with hearing loss or aphasia, wherein recipients of a third-position non-repair caused by a trouble source speaker's inability to perceive the second turn other-initiation will often choose to let the trouble pass without exerting further effort to establish understanding (Caissie & Gibson, 1997; Goodwin, 2003; Skelt, 2012). In these situations, repair is fully abandoned, as shown in Excerpt 13 and Excerpt 14 below ("C" indicates a continuation):

Excerpt 13

(Overt text bot scenario, trial 16):

Interactant: Hello, how are you?

Cleverbot: TS → No, that was wrong!

Interactant: RI → Why?

Cleverbot: NR → I want to eat a cookie.

Interactant: C → You can eat it if you want, but give me one as well.

Excerpt 14

(Covert echoborg scenario, trial 3):

Cleverbot: Why are they not funny?

Interactant: Some of them are, some of them are funny. Unintentionally funny.

Cleverbot: TS → No they're a soccer club.

Interactant: RI → What?

Cleverbot: NR → What is it like to be a human?

Interactant: C → It's weird. It's not that great.

Finally, *departures* encompassed interactant reactions to Cleverbot non-repair that neither sought to repair any preceding talk nor engaged with the content introduced by the non-repair. It featured regularly in the current study, as interactants frequently responded to Cleverbot non-repairs by simply initiating a new conversational trajectory, as demonstrated in the following excerpts (“D” indicates a departure):

Excerpt 15

(Overt echoborg scenario, trial 12):

Cleverbot: TS → So I tell him hello.

Interactant: RI → Who have you told hello?

Cleverbot: NR → Because you just asked.

Interactant: RI → A friend of yours?

Cleverbot: NR → Lol, not right. Try again.

Interactant: D → I'm thinking of going on a big trip this year. Do you have any suggestions for a vacation?

Excerpt 16

(Covert text bot scenario, trial 16):

Interactant: Do you come here often?

Cleverbot: TS → Today is a little awkward initially.

Interactant: RI → I don't understand.

Cleverbot: NR → Me too.

Interactant: D → What's your favorite film.

Table 4.1 shows how each of the four experimental conditions compare to one another in terms of interactants' responses to Cleverbot non-repairs. Although no specific hypotheses were made prior to the experiment regarding how reactions to non-repair might differ between conditions, the observation that the covert echoborg scenario (the baseline condition) had the highest relative percentage of successive repair initiations (29.14%, on average) while the overt text bot scenario saw the lowest relative percentage (22.71%, on average) supports the overarching prediction that intersubjective effort is greatest when interactants assume that

Table 4.1

Interactant Responses to Interlocutor Non-Repair

Response	Average relative percentage of response			
	Echoborg Scenario		Text Bot Scenario	
	Covert (SD)	Overt (SD)	Covert (SD)	Overt (SD)
Other-initiation	29.14% (22.81%)	26.19% (26.52%)	27.68% (24.67%)	22.72% (28.66%)
Continuation	52.29% (29.36%)	55.95% (35.95%)	62.22% (34.83%)	53.66% (37.71%)
Departure	18.57% (28.47%)	17.86% (22.09%)	10.10% (19.37%)	23.62% (30.32%)

Note. Trials: covert echoborg (n = 29), overt echoborg (n = 25), covert text bot (n = 29), overt text bot (n = 25).

they were interacting with an autonomous human and lowest when interactants knew they were communicating with an artificial agent. Lending further support to this overarching prediction is the fact that departures were most common in the overt text bot scenario (occurring in response to 23.62% of non-repairs, on average). In all conditions, interactants responded to non-repairs with continuations over 50% of the time, on average.

General Discussion

In an experimental study that made use of the echoborg method of HAI, it was demonstrated that two factors affect the likelihood of an interactant initiating repairs of misunderstandings. Participants who spoke to the chat bot Cleverbot via a text screen were less likely to utter an other-initiation of repair following a Cleverbot utterance than those who spoke to Cleverbot via a human body interface (i.e., an echoborg). Likewise, participants who were made explicitly aware prior to engaging in conversation with Cleverbot that their interlocutor's words would be determined by an agent were less likely to utter an other-initiation of repair following a Cleverbot utterance than those who were not made explicitly aware of this fact. However, results showed that the likelihood of an interactant attempting to self-repair a trouble source following an other-initiation of repair agent does not depend on whether the interactant engages with a human vs. a text screen interface or whether they are explicitly

aware of their interlocutor's words being agent-determined. A post-hoc analysis of interactant transcripts revealed that interactants were on average more likely to utter a subsequent other-initiation following a non-repair produced by their interlocutor when they both engaged with a human body and did not know their interlocutor's words to be agent-determined. Moreover, interactants more frequently departed from a repair sequence entirely when they both engaged an interlocutor through a text screen interface and knew their interlocutor's words to be agent-determined.

These results suggest that when people speak to a conversational agent under the same conditions of everyday human-human interaction (i.e., when an agent has a human body and is assumed to communicate autonomously), they more persistently try to establish common ground (i.e., exert more intersubjective effort) relative to conditions wherein knowledge that an interlocutor's words are determined by an agent is explicit and/or the interface is nonhuman. This finding is important because it points to a potential *glass ceiling* for artificial agent participation in human intersubjectivity. If roboticists were to someday design a machine that was indistinguishable from an actual human in terms of appearance and communication (i.e., if the machine were able to pass a *Total Turing Test*; see Harnad, 1991), the mere knowledge of it being something not entirely human might suppress the amount of intersubjective effort people exert when interacting with it.

Surprisingly, the likelihood of an interactant attempting to self-repair a trouble source in response to an other-initiation was not affected by the experimental manipulations whereas the rate of interactant other-initiation was. This could be because other-initiations of repair are active attempts to manipulate the attention of an interlocutor toward a trouble source and, therefore, at some level indicate an implicit supposition that an interlocutor possesses a capacity for advanced intersubjectivity. A self-repair attempt, on the other hand, is more of a reflexive response that follows a request to update an interlocutor's meta-perspective; a person need not presuppose that an interlocutor possesses a capacity for advanced intersubjectivity in order to produce a successful self-repair following other-initiation (in fact, it may even be that *not* attempting an appropriate self-repair following other-initiation, say by departing from the conversational trajectory, is more unnatural for humans than simply producing a self-repair attempt). Relative to the performance of a self-repair attempt, articulating an other-initiation of repair may involve higher-order mentalizing about the perspectives of an interlocutor, and this higher-order mentalizing may be more sensitive to

changes in how an interlocutor is represented.

These findings can be positioned within a broader discussion that concerns the centrality of intersubjectivity and intersubjective effort in human life. As argued herein, and has been established in both the fields of developmental psychology and communication, complex intersubjectivity is a co-construction (one interactant cannot accomplish it alone). Child development, for instance, requires children be brought into advanced intersubjective relations by-way-of adult scaffolding (Berk & Winsler, 1995; Plumert & Nichols-Whitehead, 1996), which entails a high level of intersubjective effort. Adults scaffold by providing verbal guidance and support for children's understanding that is *just beyond* their actual level of comprehension, thus pulling them into increasingly complex intersubjective relations (the "zone of proximal development"; see Vygotsky, 1978). Equally, achieving common ground in communicative interaction between adults cannot occur solely based on the actions or cognitions of one side of the interaction; rather it is a joint achievement, with each side supporting the other side in the calibration of perspectives (Rommetveit, 1974; Schegloff, 1992). Rommetveit's (1974) axiom "intersubjectivity has [...] to be taken for granted in order to be achieved" (p. 56) captures this notion. If humans do not initially *assume* a highly intersubjective interlocutor, they will not engage in the complex intersubjective processes with the interlocutor that are necessary to further elaborate the pair's intersubjectivity. Even when misunderstanding arises, other-initiations of repair reveal an implicit belief that common ground *can* be achieved by-way-of an exchange of perspectives. On the other hand, abandoning the assumption of intersubjectivity will block the achievement of common ground due to fewer attempts at repair.

If human interactants do not first expect high-level intersubjectivity from artificial agents, they will not extend such intersubjectivity to them, effectively locking them out of the full spectrum of human intersubjective relations. In order to develop forms of HAI that reach benchmark intersubjectivity, agents – as with human infants – will need to be able to learn from the other-initiations of repair issued by their human interactants. It is easy to imagine how severely constrained human social relations would become if interlocutors repeatedly failed to signal to each other when something has caused misunderstanding – complex joint activity would become impossible. No matter how capable each party to an interaction is at operating at benchmark intersubjectivity, it takes benchmark intersubjective effort – a robust exchanging of perspectives – to get there.

The echoborg method and analysis of repair activity as a means of benchmarking intersubjectivity in human-agent dialog. Given the primacy of intersubjectivity in cooperative human social behavior, it is imperative that evaluative frameworks for HAI generally, and human-agent dialog specifically, involve assessing intersubjectivity against human-human interaction benchmarks. Although various researchers have explicitly called for this approach (e.g., Cassell & Tartaro, 2007), numerous others have indirectly called for a focus on intersubjectivity through advocacy of interactionist HAI methodologies (e.g., Dautenhahn, 2007; Johnson et al., 2014; Parise, Kiesler, Sproull, & Waters, 1999; Payr, 2001), including but not limited to the analysis of grounding behavior in human-agent dialog (e.g., Brennan, 1991; Kiesler, 2005; Lücking & Mehler, 2014; Visser, Traum, DeVault, & op den Akker, 2014) and interaction authenticity (e.g., Feil-Seifer, Skinner, & Matarić, 2007; Kahn et al., 2007; Turkle, 2007). Indeed, a concern with intersubjectivity is implicit in many approaches to HAI evaluation, such as those that investigate people's emotional responses to robotic and virtual agents (e.g., Balzarotti, Piccini, Andreoni, & Ciceri, 2013; Brave, Nass, & Hutchinson, 2005; Prendinger & Ishizuka, 2005).

The current article is a contribution toward developing methodologies for benchmarking intersubjectivity and intersubjective effort in HAI. The current article demonstrates how the echoborg method can be used to evaluate human-agent intersubjectivity when the agent is an artificial conversational agent. The unique strength of the echoborg method is that it can involve research participants communicating with an agent *while believing that they are speaking to an autonomous human being*. Thus, the echoborg method allows researchers the ability to investigate HAI intersubjectivity while preserving the interactant's sense that they are experiencing a fully human-human social psychological context. Future research can evaluate nonhuman interfaces by comparing intersubjective effort with a baseline echoborg condition.

Might intersubjective effort in human-agent dialog increase as technology improves? The seeming unwillingness of humans to exert benchmark intersubjective effort with a conversational agent in the non-baseline conditions of the present research may prove to be a historical artifact. As more advanced artificial intelligence develops and as people are raised in a world in which socially advanced artificial agents are ubiquitous, the expectations people will place on the intersubjective capacities of their machine interlocutors may increase. The study of social psychological phenomena is, in many respects, the study of behavioral

patterns that are contingent upon cultural and historical circumstances (Gergen, 1973). The more that artificial agents are able to engage humans in complex intersubjective processes, such as repair work, the more humans will scaffold their participation in such intersubjectivity. Indeed, the finding that interactants self-repair at relatively consistent rates irrespective of the interface they engage with or their awareness of the agency of an interlocutor is evidence that humans readily “step up” and respond with a repair attempt as needed.

Limitations. A particular strength of the echoborg method, which itself is derived from the “cyranoid method” of social interaction (Corti & Gillespie, 2015b), is that it enables the study of social interactions that are high in mundane realism (dynamic, unscripted and face-to-face). The benefits of such realism, however, incur costs in the form of certain control limitations. For instance, though best efforts were made to ensure that the speech shadower’s body language was consistent across experimental trials (the shadower was instructed to try and match their body language to the words they found themselves articulating), it is all but impossible to completely eliminate variability in shadower body language from trial to trial using the echoborg method. Furthermore, the minimal technological dependency format of inputting the interactant’s speech into Cleverbot creates slightly unnatural delays between interactant utterances and subsequent responses by the agent. Though all experimental conditions were subject to the same latencies, future improvements to the echoborg method may mitigate these limitations.

Conclusion. The present research has found that two factors significantly affect the rate at which an interactant will attempt to elicit repairs of misunderstandings from an interlocutor that is a conversational agent, namely, (1) the agent’s interface and (2) the interactant’s awareness that their interlocutor’s words are agent-determined. However, these factors did not affect the rate at which interactants attempted to self-repair misunderstandings following a request. It seems that people exert the most intersubjective effort with an agent (operationalized as repair initiation) when they are unaware that their interlocutor’s words are determined by an agent and when the agent interfaces via an actual human body (i.e., an echoborg).

This article contributes a novel methodology (i.e., the echoborg method) to the study of HAI and demonstrates how it can be used to compare conditions of HAI that make use of

nonhuman interfaces and nonhuman agency-framing to a baseline condition that approximates the social psychological contextual frame experience by people during mundane, face-to-face, in person, human-human interaction. The findings of the present study have implications for the development of autonomous social agents. Most notably, if agents are to participate fully in the intersubjective world of humans, not only must they be *capable* of interacting at benchmark intersubjectivity, but human interactants must also be willing to exert intersubjective effort at a level conducive for the achievement of benchmark intersubjectivity (e.g., by uttering other-initiations of repair when misunderstanding arises). Artificial agents cannot enter the world of human intersubjectivity without the support of their human interactants, and this support is contingent upon the human interactants' supposition that complex intersubjectivity is achievable.

CHAPTER 5 | Cyranoids in First-Person, Self-Experimental Research

Preface

Each of the studies so far described in this thesis has primarily focused on the behaviors, attributions, and perceptions of interactants who have come into contact with a cyranoid. This chapter turns attention toward the other two actors involved in a cyranic interaction, the source and the shadower, as no exploration of the cyranoid method would be complete without considering the experiences of the individuals who actually comprise a cyranoid. The source within a cyranoid is faced with speaking through someone else's body in real-time. Thus, they engage in real-world social interactions with a transformed outer identity that shapes how they are perceived and reacted to by interactants. Shadowers, on the other hand, experience social interactions with a transformed inner identity, or disposition, and this likewise shapes how they are perceived and reacted to by interactants. In the case of the echoborg, the human shadower's inner identity is transformed so as to become that of an artificial conversational agent, and as a consequence, echoborg shadowers have the embodied experience of having their verbal agency controlled, partially or entirely, by a machine.

These affordances of the cyranoid method present opportunities to study how transformations (or more specifically, hybridizations) of outer and inner identity make an impact upon those who partake in such transformations. One way to study this impact would be to assign research participants to either the role of a source or a shadower and gather data accordingly. Alternatively, a researcher can place themselves in the position of the source or the shadower and engage in a research context in a direct, participatory fashion, and in doing so gain experiential knowledge as to the effects identity hybridization brings to bear in a particular social context. Via the cyranoid method, one can step into the shoes, so to speak, of another human being, and gain a first-person sense of how the world that person lives in responds to parts of their outer or inner identity. Moreover, the cyranoid method allows one to assume such a transformed identity in repeated and controlled trials, meaning that one can build robust experiential knowledge of the social consequences of a particular aspect of identity iteratively by designing a cyranic manipulation and conducting self-experimentation. This harkens back to Popper's (1977, 1979) and Green's (1983) ideas discussed in Chapter 1 concerning how individuals can gain improved self-knowledge via directly manipulating some aspect of World 1 self-embodiment and observing the associated affects.

Arguing in favor of a form of researcher-as-subject, self-experimentation, however, runs counter a status quo within contemporary experimental social psychology that sees such types of research as anathema to certain methodological principles. This mainstream position, however, is largely an artifact born of historical forces within the discipline that saw first-person methods largely split from experimental psychology in the early/mid-20th century for what at the time were largely institutional reasons. A first-person, researcher-as-subject tradition remains alive and well, however, in branches of the human sciences adjacent to social psychology, particularly those disciplines that make frequent use of field methodologies such as participant observation and autoethnography. Milgram, having come from an intellectual tradition within social psychology that valued first-person methodology and direct experience, was, during his time, among a relatively small cohort of experimental social psychologists that practiced first-person methods, as is evident in both his cyranoid pilot studies (he frequently functioned as the source or shadower in his pilots and reported first-person reflections from these experiences; see Milgram, 2010a) as well as in his breaching experiment field studies (e.g., Milgram 2010c, 2010d). Much of this chapter is devoted to retracing the rise and fall of self-experimentation within psychology, discussing where within the discipline self-experimentation has re-emerged in recent years, and discussing how the cyranoid method can contribute as a framework for researchers wishing to self-experiment with transformations of identity in social contexts.

The journal article that follows was developed following discussions with other social psychologists who have designed and/or participated in cyranoid experiments as either a source or a shadower and documented their own experience. The article presents their first-person experiences in these roles and the social psychological insights they garnered. Three potential benefits of this type of researcher-as-subject self-experimentation are suggested: (1) direct access to “social qualia,” or the subjective experiential knowledge of real-world social phenomena, (2) an improvement to mental models concerning social psychological phenomena, which in turn can stimulate new research questions and insights, and (3) enhanced reflexivity about a given experiment (namely, the embodied understanding of how one’s outer or inner identity plays a direct role in shaping how the social world responds to the self’s activity within it. Thus, researcher-as-subject self-experimentation is not put forward as a substitute for traditional, third-person experimentation (in which the researcher is separate from the subjects), but as a research practice that can enhance the perspectival knowledge of certain phenomena, for a first-person *and* a third-person understanding of a

given social psychological phenomenon (e.g., stereotyping) is certainly more beneficial to the researcher than either of these perspectives alone.

The journal article was authored by Corti, Reddy, Choi and Gillespie (2015). Corti outlined the article, authored the main drafts, and provided a first-person self-reflection, contributing roughly 75% of the content. Reddy and Choi each provided a first-person self-reflection. Gillespie provided key supervisory assistance, provided Corti with suggestions for the article, and contributed roughly 25% of the content. The editor of *Integrative Psychological and Behavioral Science*, as well as two anonymous reviewers, provided helpful feedback, suggested edits, and ultimately approved the article for publication.

CHAPTER 5 | Cyranoids in First-Person, Self-Experimental Research

Article Title

The researcher as experimental subject: Using self-experimentation to access experiences, understand social phenomena, and stimulate reflexivity

Authors

Kevin Corti (primary)

Geetha Reddy

Ellen Choi

Alex Gillespie

Abstract

The current article argues that researcher-as-subject self-experimentation can provide valuable insight and systematic knowledge to social psychologists. This approach, the modus operandi of experimental psychology when the field was in its infancy, has been largely eclipsed by an almost exclusive focus on participant-as-subject other-experimentation. Drawing from the non-experimental first-person traditions of autoethnography, participant observation, and phenomenology, we argue that participating as both observer and subject within one's own social psychological experiment affords researchers at least three potential benefits: (1) access to "social qualia," that is, the subjective experience of social phenomena; (2) improved mental models of social phenomena, potentially stimulating new research questions; and (3) an enhanced ability to be reflexive about the given experiment. To support our position, we provide first-person self- reflections from researchers who have self-experimented with transformed social interactions involving Milgram's cyranoid method. We close by offering guidelines on how one might approach self-experimentation, and discuss a variety of first-person perspective ethnographic technologies that can be incorporated into the practice.

Keywords

Cyranoid, first-person methodology, Milgram, phenomenology, self-experimentation

Introduction

What counts as an experiment? The question has been asked by Winston and Blais (1996),

whose analysis of textbook guidelines on experimental methods from an array of disciplines found that the field of psychology has come to adopt a uniquely rigid, narrow, and uniform stance regarding how experiments are defined. They attribute this phenomenon in part to a belief which emerged during the 20th century among psychologists that third-person (i.e., participant-as-subject) controlled experimentation could produce knowledge beyond that provided by disciplines that make use of non-controlled or first-person methodologies, and that this knowledge could point toward means of reshaping individual and social behavior. This development has had the laudable aim of making experimental psychology rigorously scientific; however, it is a moot point whether a narrowing of focus and an exclusion of relevant data is indeed scientific. A consequence of an almost exclusive focus on third-person methodology has been the disappearance of a researcher-as-subject, first-person perspective approach to social psychological experimentation.

Any examination of the vast contemporary experimental social psychology literature will reveal the methodological orthodoxy regarding the role of the researcher vis-à-vis those who participate in research. The researcher is positioned as the expert observer who, from an ostensibly independent third-person position, evaluates a naïve subject who has been exposed to carefully operationalized stimuli. The range of subjective experiences of both the researcher and the subject are typically of secondary or negligible significance to the aims of the experiment as primary focus is placed on measuring a narrow set of activities or self-report items corresponding to pre-determined hypotheses. Upon uniformly gathering data from a critical mass of subjects, the researcher performs and reports the results of statistical inference tests to validate, refine, or challenge theory - theories which themselves are usually defined from a third-person point of view in terms of predicting the behavior of others.

So predominant is this basic template of experimentation within the discipline that it is hard to imagine experimental practice taking a different form. We argue that a different form is not only possible, but also evident in the origins of experimental psychology as an independent field of inquiry, wherein the laboratory approach often involved direct interaction among researchers and their subjects as well as a much more egalitarian and interchangeable relationship between these roles. Indeed, experimental psychology began as a discipline of researcher-as-subject self-experimentation. Though this orientation was largely abandoned in the early 20th century, traces of the approach have survived in genres of psychology closely aligned with anthropology and sociology, fields in which a non-experimental researcher-as-

subject investigatory model is alive and well. Arguably, it also persists in experimental social psychology itself, feeding into the design and development of experiments, but it is rarely reported and given the legitimacy that it deserves. In the current paper, we revisit a particular technique of first-person self-experimentation developed in the late 1970s by the social psychologist Stanley Milgram (his “cyranoid method”) and vet its utility as a tool for informing researchers as to certain experiential elements of social phenomena.

We should make it unambiguously clear at the outset of this endeavor that our purpose herein is not to challenge the position or validity of today’s mainstream third-person approach to social psychological experimentation. We do not suggest that self-experimentation should supplant other-experimentation, but rather, that it might augment and enrich other-experimentation. We propose that the experimental method becomes impoverished by excluding the rich, lived-in, subjective first-person experience of research subjects from consideration and by making researcher-subject role segregation overly rigid. The question is whether or not self-experimentation has anything to offer as an additional methodology, and whether or not such a methodology could uniquely service certain aims of social psychological researchers, namely those not sufficed by a third-person approach. Such open-ended questions can only be settled through a renewed discussion and reappraisal of the merits of self-experimentation (and a researcher-as-subject model more generally), and the current work is our initial contribution to such a reappraisal.

We feel strongly that revived focus on self-experimentation is warranted given how ethnographic practices, such as autoethnography and participant-observation (both in the real world and in virtual worlds), have evolved in recent years. Indeed, recent research provides many illustrations of how new technologies and digital methodologies enable forms of first-person perspective research (e.g., Lahlou, Le Bellu, and Boesen-Mariani 2015) that were unavailable (and perhaps unimaginable) to experimenters of prior generations, yet among whom many of the issues regarding the legitimacy of first-person self-observation methods were supposedly settled. Further impetus for exploring new formats of self-experimentation in social psychology is provided by recent debate concerning the status of phenomenology (and its primary method, introspection) in relation to psychology (see Dennett 2007; Marbach 2007), the details of which we will touch upon briefly. In our discussion we outline a number of procedural proposals for researchers interested in deploying first-person methodologies for self-experimentation, and argue that participating in one’s own social experiment and being

exposed to experimental manipulations first-hand affords researchers at least three potential benefits: (1) the acquisition of subjective experiential knowledge (i.e., “social qualia”) regarding particular social psychological phenomena that could otherwise only be understood descriptively or indirectly through third-person perspective analysis; (2) the development of richer mental models regarding the nature of social phenomena in the world beyond the experimental setting, which in turn can stimulate new research questions; and (3) the improved ability to be reflexive about an experiment by virtue of understanding the co-occurring perspectives within the experimental setting.

A Brief History of Psychological Self-Experimentation

In *Constructing the Subject: Historical Origins of Psychological Research*, Danziger (1990) traces the development and eventual eclipse of the researcher-as-subject model of psychological experimentation (also see Danziger 1980, 1985; Farr 1996). He begins by discussing the first formal models of experimental psychology to emerge in the 19th century, the predominant being Wilhelm Wundt’s Leipzig model and the Paris model. Wundt’s model consisted of academic researchers, colleagues, and students engaging in a highly collaborative research enterprise in which self-observation methods were at the fore. Central to this model was Wundt’s assertion that the ideal experimental subject was an expert observer - a psychologist trained in methods of self-perception capable of carrying out experiments on his or her own self (Kusch 1995). Individuals involved in studying a particular psychological phenomenon would often exchange roles in an experiment, functioning at times as the experimenter and at other times as the subject (i.e., the expert observer), while the nature of phenomena would be deduced via collaborative analysis. However, alternative approaches being developed in Paris and London did not have any role or position exchange between subjects and researchers. The Paris model developed from the clinical assessment of patients in medical contexts and featured strict role segregation between researchers and subjects, with subjects serving as objects of study by expert clinicians. Meanwhile, Francis Galton’s London laboratory entailed psychologists providing mental faculty testing services to members of the public in return for a fee, with strict division between the experts and lay people. Danziger (1990) notes how Galton’s approach introduced the “multiplication of subjects” (p. 57) to psychology: as it was necessary for individual cases to be validly compared with one another, statistical inference was required, thus necessitating both the transposition of mental phenomena into quantitative data as well as large sample sizes. The case of Galton is interesting, because he was also a systematic user

of self-experimentation in his pioneering work on associations to places and words (Galton 1879); he thus demonstrates how, in the early years of psychology, both self-experimentation and other-experimentation sat comfortably side-by-side.

Psychologists in the Würzburg school argued that Wundt's methods should be extended beyond rudimentary mental imagery into an exploration of the higher mental functions (Mischel 1970; Wagoner 2009). They sought to examine how thoughts come into being and how they change. Würzburg experimenters elicited qualitative accounts, and, recognizing that introspection could alter psychological processes, also made use of retrospection (Danziger 1980). Broadening the introspective tool-kit enabled the Würzburg school to generate findings that challenged Wundt's simplistic accounts of psychological processes (Gillespie and Zittoun 2010). At the same time, Edward Titchener, a student of Wundt, founded a large and highly influential structuralist psychology program at Cornell University and sought to establish his strict version of introspection as the indispensable experimental method for investigating higher mental functions (see Titchener 1927). Despite the schisms among the early self-observationalists, there was common appreciation for a laboratory model in which the observer and subject were one and the same person (e.g., see Edwin Boring's self-experiments on sensory experience: Boring 1915), and the researcher-as-subject approach enjoyed success in both North America and Europe during experimental psychology's formative years (Vermersch 1999).

While Wundt's methods were criticized as being not subjective enough by the alternative schools of self-observation, they were criticized by others as being too subjective. This was the positivist repudiation of Wundt (Danziger 1979). Psychologists such as Ebbinghaus and subsequently the behaviorists argued that Wundt's concepts were too metaphysical and, consequently, not sufficiently observable. They advocated stripping down the method and focusing on more objectively observable phenomena such as behavior. This intellectual shift coincided with practical demands in the United States for psychologists to deliver practical scientific findings to consumers of psychological research (e.g., government). Thus, there began a "fundamental shift of interest from the analysis of psychological processes, necessarily manifested in specific individuals, to the distribution of psychological characteristics in populations [...] what emerged was an impersonal style of research in which experimental subjects played an anonymous role, experimenter-subject contacts were relatively brief, and the experimenter was interested in the aggregate data to be obtained from

many subjects” (Danziger 1985, p. 137). As an embrace of logical positivism throughout the natural sciences brought behaviorism to the fore of experimental psychology, self-observation was deemed by many as an inappropriate approach to laboratory science (Danziger 1979, 2000; Farr 1983). As Farr (1978, p. 302) explains, “the experimenter now became the ‘observer’ and it was the behavior of the subject rather than his experience which constituted the raw data.”

While behaviorism’s third-person approach to experimentation supplanted the first-person model throughout much of psychology, there were notable exceptions. Specifically, a tradition of inquiry within social psychology grew out of Husserl’s phenomenology, which was carried to the United States by Alfred Schutz and Gustav Ichheiser and inspired the thinking of people such as Berger and Luckmann (1967), Goffman (1959), and Garfinkel (1967). Consider Harold Garfinkel’s (1967) “breaching experiments,” wherein researchers would consciously exhibit a contextually inappropriate behavior in a mundane social situation in order to both observe the social reactions elicited by such behavior as well as (and more relevant to the notion of self-study) gain experiential insight into the nature of violations of social expectations (Rafalovich 2006). Stanley Milgram, famous for his controversial studies on obedience to authority (Milgram 1963), adopted Garfinkel’s approach in several field experiments. These included a study into whether random strangers on New York City subways would give up their seats upon request to a covert experimenter, and while the study primarily focused on the reaction and compliance of the subway riders party to these requests, Milgram had his experimenters record and reflect on their own emotions, behavior, and psychological states during the interactions (Milgram 2010b). In another field experiment, Milgram’s cohort of experimenters systematically intruded into waiting lines and reported on the range of emotions felt prior to, during, and after these norm-breaching events, thus directly experiencing the “inhibitory anxiety that ordinarily prevents individuals from breaching social norms” (Milgram 2010c, pp. 49–50).

By the late 1970s, however, social psychology too began to more forcefully shed itself of the researcher-as-subject orientation as well as experimental approaches that featured high degrees of researcher-participant interaction as the field passed through what is now considered an era of “crisis” (see Adair 1991; Elms 1975). Under pressure to defend the discipline’s position among the natural sciences at a time when experimental findings risked being exposed as non-replicable or otherwise untenable, researchers disavowed methodology

deemed overly susceptible to intrusion from latent, uncontrolled phenomena (e.g., experimenter effects; Rosenthal 1966). To counter the possibility of their findings being challenged as mere artifacts of experimental idiosyncrasy, social psychologists sought to sanitize their laboratories and experimental procedures of any and all potential confounds. This evolution stripped social experiments of their mundane realism, atomized dependent variables (Farr 1978), and further solidified the notion that the first-person experiences of researchers were inappropriate objects of experimental evaluation. The modern experimental model which emerged from this era has been successful for cognitive third-person approaches to social psychology (Adair 1991), but has led to an entrenched methodological status-quo generally unreceptive to pre-crisis (and Milgram-esque) methodology (see Shotter 2015).

Contemporary Researcher-as-Subject Approaches

While a researcher-as-subject approach in social psychology has largely dissipated, one can still find traces of it in various corners of psychological science. Phenomenological psychology, for instance, lived for many decades far outside mainstream psychology producing Husserlian self-studies on the nature of experiential phenomena, yet has in the last decade attracted increased attention and elicited loud debate concerning whether an objective approach to understanding the mind can (and should) involve first-person research (for discussions on the historical trajectory of 20th century phenomenological psychology, see Giorgi 1998; Klein and Westcott 1994). Some proponents vigorously defend the practice and legitimacy of autophenomenology (e.g., Marbach 2007; Varela and Shear 1999) in contrast to others who advocate a more guarded approach that seeks to verify first-person experience via third-person data (Dennett's heterophenomenology; Dennett 2007), while others point out that variants of introspective reports (e.g., self-report questionnaires) are ubiquitous throughout psychology as it is, and that the domains of emotion, attitude, memory, and developmental research attest to this fact (Wilson 2003). Despite this renewed focus on phenomenology within psychology, phenomenological studies primarily involve subjects reporting to a researcher their beliefs about the conscious phenomena they experience in a given experimental condition, and most of the studies reported in the literature involve not social experiences per se, but basic perceptual experiences that speak more to researchers in cognitive science than to social researchers (for examples of typical contemporary phenomenological experiments, see Gallagher and Sørensen 2006). Therefore, contemporary phenomenological psychology does not so much provide us with a template for a researcher-as-subject model of social experimentation as provide us with an example of a modern

discipline that takes seriously the value in capturing experiential phenomena through first-person experimental procedures.

Analogues to researcher-as-subject experimental social psychology have thrived in the non-experimental anthropological and sociological traditions of autoethnography and participant observation. Autoethnography is a self-referential form of qualitative analysis wherein a researcher becomes embedded in a particular social context and reports on the scope of his or her subjective experiences and self-transformations (Anderson 2006). Thus, the primary data of autoethnographers are autobiographical first-person self-reflections (Chang 2008). Participant observation, meanwhile, is a field methodology general to ethnography wherein the researcher actively assumes a participatory role with others in a given social frame, either overtly as a known-researcher or covertly in a “disguised role” (Becker and Geer 1957). Indeed, one of the originators of participant observation, Bronislaw Malinowski, was heavily influenced by Wundt while a student at Leipzig (Farr 1983; Strenski 1982). While neither autoethnography nor participant observation necessarily seek to make claims regarding phenomenological causality, each adopts a first-person perspective stance that is similar to that found in the early self-observation experimental research. Autoethnography, with its researcher-as-subject orientation, embraces the value in experiential and self-reflexive knowledge, while participant observation, being rooted in interactivity, dissolves the traditional researcher-subject role boundaries common to modern experimentation and allows for direct interaction among researchers and their participants.

Some researchers within contemporary behaviorism advocate for a researcher-as-subject self-experimental approach. Neuringer (1981), for instance, outlines a number of potential advantages for experimental psychologists who practice and report self-experimentation, namely that it would catalyze experimental process discoveries as well as foster a heightened experimental ethic among communities of researchers. In a series of self-reflections on the benefits of long-term psychological self-experimentation, Roberts (2004, 2012) argues that in addition to allowing the researcher to test new concepts cheaply, positioning oneself as both experimenter and subject generates the types of experiential knowledge from which one may formulate new research ideas distinct from those that arise through synthesizing empirical literature. Much of this advocacy points to the long-standing tradition of self-experimentation in the medical sciences, wherein vital breakthroughs have emerged through self-study (e.g., Nobel Prize recipient Barry Marshall’s controlled ingestion of *H. pylori* bacteria which

revealed its culpability in causing gastritis). While the nature of researcher-as-subject self-experimentation of the sort we're advocating does not involve subjecting oneself to potential harm, the example of the medical tradition as well as the modern calls from within behaviorism (the movement which, after all, was instrumental in ridding psychology of its first-person orientation nearly a century ago) suggest that perhaps it is time for a larger discussion within social psychology regarding the possibility of a contemporary researcher-as-subject model of experimentation.

Milgram's Cyranoid Method

Our own interest in self-experimentation began by delving into the life and work of Stanley Milgram, who as the field experiments described above suggest, oriented to a phenomenological approach to social psychology (see Zimbardo 1992). He is known for his ground-breaking research, his highly exploratory methods of experimentation, as well as his penchant for participating in his own studies. Milgram's (2010b, 2010c) breaching experiments attest to the fact that his desire to understand the social psychology of others was equally matched by his desire for he and his fellow researchers to learn about themselves through the process of experimentation. In this regard, his cyranoid studies - conducted in the years leading up to his death, and never formally reported - provide one of the more striking examples of his first-person approach.

Milgram's cyranoid studies involved first training research confederates to speech shadow (i.e., replicate the speech of a third-party in near real-time; see Marslen-Wilson 1973), and subsequently, by-way-of covert radio relay, conversing through these shadowers with research participants who were unaware their conversation partner was merely replicating Milgram's words (Milgram 2010a; Blass 2004). Milgram's participants repeatedly failed to detect that their interlocutor was a speech shadower, seemingly taking for granted the verbal autonomy of the person with whom they interacted (a phenomenon referred to as the "cyranic illusion"). He called the hybrid agent composed of his mind (or rather, words) and a shadower's body a "cyranoid," a term that paid homage to the character Cyrano from the play *Cyrano de Bergerac*.

The theoretical thrust of the cyranoid studies involves questions regarding the relationship between person perception and the subjective experience of self. Milgram (2010a) felt that with the method people could experience "radical deformations" (p. 408) of self during social

interactions mediated by a shadower's body that was vastly divergent from their own in terms of gender, ethnicity, age, social status, and so on. At the time of Milgram's cyranoid studies it was well known that people stereotype other individuals on the basis of their outward identity (e.g., Cantor and Mischel 1979; Tajfel, Sheikh, and Gardner 1964) and that people's behavior in many ways confirms the stereotypes held by others (e.g., Snyder, Tanke, and Berscheid 1977), but experimental validations of these phenomena were largely, if not entirely, third-person in nature. So whereas the traditional literature showed that people stereotype (along with descriptions on the nature of stereotyping in various contexts), Milgram invented an experimental technique whereby one could systematically manipulate the experience of being stereotyped during social interactions in accordance with the outer identity of an interchangeable shadower.

In his most elaborate iteration of the cyranoid method, Milgram (2010a) separately sourced for 11- and 12-year-old children while being interviewed by panels of teachers tasked with assessing their interviewee's intellectual capabilities (for a replication of this study, see Corti and Gillespie 2015). The teachers were unaware their interlocutors were in fact shadowers articulating the words of a college professor, thus Milgram experienced conversing with these teachers as though he had the identity of a child. In his self-reflections following these interactions, Milgram noted that despite his best efforts to impress the teachers with his knowledge (he was, after all, a Harvard-educated professor), the outer persona of his child shadowers restricted to a large degree the manner in which teachers engaged with him. As the teachers believed they were speaking to a child, they naturally addressed their interviewees as children despite the sophisticated responses Milgram produced, leaving him unable to signal the full scope of his intelligence to the participants. His sole reportage of these studies is very autoethnographic in tone and comes in the form of a speech he prepared for an American Psychological Association convention in Toronto in 1984, an excerpt from which reads:

The technique is an extraordinary prism through which to understand how people form judgments of others. For in a significant degree, the opinions of teachers formed of our child cyranoids depended as much on the teacher as on the child, and the questions asked and avoided. Teachers varied in how they approached questions [...] the worst never seeing beyond the possibilities of an average 11-year-old [...] Often, teachers

themselves simply did not have the knowledge, information, or inclination to ask adequate questions[...] We thus see very clearly how the impressions people form are to some extent generated by their own interaction with the stimulus person, the things they bring out and suppress (Milgram 2010a, p. 407).

A Method for Self-Experimenting with Altered Identities. The interchangeable componentry of the cyranoid method provides a highly versatile means for researchers to explore transformed social interactions. There are three parts to a cyranic interaction: (1) the source (the person whose mind provides words to the shadower), (2) the shadower (the person whose body reproduces the source's words), and (3) the interactant (the person who converses with the shadower face-to-face and who may either be aware or unaware that their interlocutor is a cyranoid). A researcher may assume any of these three positions. Milgram's (2010a) studies are an example of a researcher experiencing a social encounter with an alternate outer identity in that he sourced for differentiated shadower-types. However, one may also take the position of the shadower and in doing so experience social encounters wherein the words one contributes are not one's own, but that of a remote source. A researcher may also experience socializing from the perspective of the interactant and encounter cyranoids whose outer identity is readily perceivable but whose inner identity (that of the source) is not. Furthermore, multiple sources and/or shadowers may be used. For example, one source may speak through two or more shadowers and in doing so experience what it is like to simultaneously be different people in a social setting; likewise, two or more sources can control a single shadower, creating a situation whereby multiple individuals are projected through the same body.

In addition to their unique phenomenological qualities, each of the three positions within a cyranic interaction can be distinguished in terms of the research questions they raise. For example, if one wishes, as Milgram had, to experience being perceived as someone quite dissimilar from one's real self, one may take the position of the source and self-experiment with different varieties of shadowers. Such scenarios raise classic questions about the nature of identity and social perception (e.g., how a source's behavior may confirm stereotyped assumptions about their shadower; see "the Proteus effect": Yee and Bailenson 2007). On the other hand, by assuming the role of a shadower a researcher can investigate, for example, how their inner cognitions and sensations vary in accordance with the words they shadow,

which may diverge from their own communicative tendency in terms of structure, idiosyncrasy, ideology, cultural disposition, and so on. Finally, by taking the perspective of an interactant, a researcher may self-experiment with issues such as heuristic and systematic communicative information processing; for instance, one may observe oneself gauging the overt characteristics of an interlocutor (e.g., their physical appearance) in relation to the informational value in what they speak.

Cyranoid Analogues in Virtual Ethnography and Autoethnography. We can connect the cyranoid method to contemporary ethnographic and autoethnographic research on the experience of self in transformative contexts. Milgram's (2010a) fascination with cyranoids arose from considering the "psychological consequences" of a "hypothetical world in which the thoughts of one person would come out of another person's mouth" (p. 402), and though such a world in a literal sense does not exist (today), online virtual worlds such as the immersive 3-D platform Second Life serve as perhaps the closest conceptual analogues. In these communities, users socialize through avatars: digital self-representations that may or may not mirror the self of reality. Avatars allow users to engage in what Nakamura (2002) calls "identity tourism," wherein one can experience the world through the eyes of whatever persona they wish to concoct. The last two decades have seen growing concentrations of literature from researchers who have purposely augmented their identity in online worlds in order to examine the relationship between first-person perception, self-reflection, and the co-construction of social phenomena (for an overview, see Williams 2007). For instance, in the context of an ethnographic study of a Second Life community, Boellstorff (2008) describes how "newbies" (a term for users with limited/superficial familiarity with Second Life) are identified on the basis of whether or not their avatar's physical characteristics resemble the default settings of the software (thus suggesting restricted understanding of the more refined aspects of the platform), and subsequently how this identifier in many respects shaped the nature of his interactions with other virtual users. In an autoethnographic account, Dumitrica and Gaden (2009) describe the nature of gender-role construction and reproduction within Second Life and reflect on the experience of embodying various physically-differentiated avatars while developing a first-person sense of the social norms and roles that were subsequently expected of them by other users in accordance with these outer identities.

Virtual self-studies of this kind have been demonstrated as a reflexive practice. Gottschalk (2010) states that "we must both be self-reflexive about how we represent ourselves in virtual

spaces and keep in mind that we can choose to represent ourselves in ways that will facilitate access, entrée, and rapport" (p. 520). Reflexive cogitations, according to Ikegami (2011), occur quite readily during avatar-embodiment, as an awareness of one's ability inhabit a plurality of outer personae leads to reflections regarding the extent to which the self and its intersubjective experiences are shaped by outer appearance. She invokes Young and Whitty's (2010) maxim regarding cyberspace and embodiment: "the more we try to disengage with the body, the more its importance is revealed to us" (p. 209). By extension, we can see how the cyranoid method can serve as a tremendous resource to self-experimenters: it allows us to systematically step outside of ourselves and into a new identity, and thereby experientially gain knowledge as to how outer identity is connected to the social phenomena we hope to understand.

To showcase the quality of unique insights this method provides, the following section provides three first-person accounts from researchers who self-experimented with the cyranoid method in different social contexts.

Reflections from Researchers Who Have Self-Experimented with Cyranoids

Participating in Decision-Making Groups via Age-Differentiated Shadowers. The study undertaken by KC (first author) concerned the experience of having to argue before a decision-making group in favor of a counter-intuitive strategy related to a resource-prioritization task (the Subarctic Survival Situation; Lafferty 1987) through covert shadowers differentiated by age-group. At issue was whether or not these experiences would differ on account of whether or not the shadower's age-group, and therefore the researcher's perceived outer identity, aligned with that of the other group members.

There were two conditions in the study, a Young-adult condition in which KC (a Caucasian male, aged 25) sourced for a young-adult shadower (Caucasian male, aged 22) in four separate discussion groups, each composed of between 4 to 7 young-adult research participants (mean age = 24.2), and a Middle-aged condition in which KC sourced for a middle-aged shadower (Caucasian male, aged 44) in four separate discussion groups, each composed of between 4 to 8 young-adult research participants (mean age = 23.8). Research participants were recruited from a major North American city via internet advertisement. The conditions were designed such that in the Middle-aged condition the shadower was clearly the most senior member of each group in terms of age, whereas in the Young-adult condition

the shadower was roughly the same age as other group members. Though research participants were aware their discussions were being observed remotely by the researcher, none were aware until a debriefing session following the interactions that the researcher was in fact communicating through a covert shadower as a confederate group member. Each group discussed solutions to the resource-prioritization task for 15 min, during which KC, through the body of a shadower, argued in favor of an objectively terrible strategy (according to Lafferty 1987). KC, who was positioned in a sound-proof room adjacent to the discussion, observed groups via an audio-visual feed and provided words via a covert radio apparatus that transmitted to a discreet inner-ear radio receiver worn by the shadower.

Reflection. I entered this study with no firm expectations as to how my experiences would differ depending on the identity of the two shadowers, but it did not take long to notice differences. Though in reality I was in the same age cohort as the participants in my study, I found myself having to work harder to convince the decision-making groups of my strategy when I communicated through the middle-aged shadower compared to when I communicated through the young-adult shadower. This boiled down to the fact that participants in the middle-aged condition asked of me many more questions regarding the strategy I advocated, and therefore I had to generate many more on-the-spot rationalizations for my position. Even though the strategy I was advocating was designed to be terrible, I received many more supportive comments from the participants when I sourced for the middle-aged shadower, a pattern I attributed to the fact that in these conditions I was seen as far-and-away the most senior member of the group in terms of age, and this somehow made me worthy of more flattery than the strategy I was espousing deserved.

The experience has certainly illuminated for me the role of outer identity in mediating social experience. Far from just a narrow range of phenomena changing depending on the body

of my shadower, an array of nuances was subtly impacted. For instance, as fewer questions were asked of me when I sourced for the young-adult shadower, I felt a strong urge to speak for longer and with more emphasis when a question or comment was directed to me in comparison to when I sourced for the middle-aged shadower. In other words, as opportunities to exert influence were rarer in the Young-adult condition, I felt I had to make the opportunities count for more, the effect of which was that my overall tone in the Young-adult condition was somewhat more serious and my statements were more measured. This has made me more reflective of the intimate connection between identity and social perception, and how joint activities can traverse an entirely different trajectory depending on one's outer appearance.

The knowledge I took from the study includes a more acute awareness of how the subtle aspects of stereotyping shape social interactions, even if in an innocuous way. Having age-seniority in the Middle-aged condition discussion groups allowed me the time and space to articulate my ideas, and consequently, opportunities to exert influence (for better or for worse). Perhaps in real world contexts being given an opportunity to speak on the basis of one's supposed identity can inflate a person's sense of importance and/or the value of their ideas in the eyes of other group members. (KC, first author).

Experiencing Job Interviews as Both a Male and Female Applicant. EC's (third author) study took place in a Canadian branch of a publicly traded multinational organization. The cyranoid method was employed to allow the researcher to experience the dynamics and social exchanges of personnel selection interviews through the body of different applicants (shadowers), and consequently, the hiring decisions. 20 employees were recruited as participant-interviewers and randomly assigned to one of two conditions, a Male-applicant condition (in which the shadower was a 26-year-old Caucasian male actor) and a Female-applicant condition (in which the shadower was a 29-year-old female fashion model of mixed

Asian and Caucasian descent). When selecting shadowers, best attempts were made to control for relative physical attractiveness, and cyranoids were trained so that each would have consistent body language across experimental trials. This was so that any variability in experience between the two conditions could be attributed to the gender difference between the shadowers. EC was then separately interviewed by the participant-interviewers (90 % of whom were male) in trials that alternated by condition, resulting in 10 interviews and evaluations for both the Male-applicant and Female-applicant conditions. During the interviews, EC watched over a live audio-visual feed from an adjacent room and transmitted responses to the shadower via a hidden earpiece (in a manner similar to that employed in the aforementioned study).

Employees were told that they were participating in a research study on decision-making and that they should make best attempts to evaluate whether the applicant was suitable to hire. After each trial, participants were made aware of the cyranic nature of the interactions in a debriefing session. In the end, the applicant in Female-applicant was recommend for hire 9 out of 10 times versus 6 out of 10 times for Male-applicant.

Reflection. It was, in short, an astonishing experience. The body of the candidate markedly influenced the social dynamic within the interview and to experience these differences first-hand was thrilling. These interviews were unlike any interviews I had been in myself, suggesting that there were forces beyond gender alone at play. First, let me describe the experience of sourcing for the female shadower. While I myself am female, I am not a fashion model. In the body of the female shadower, I often felt that my interviewers were gazing at me, sometimes to the point of flirtation. They treated me very politely, perhaps like men are ‘supposed’ to treat ladies. I felt myself responding in kind and muting or exaggerating my personality to resonate most with how the interviewers saw me. Overall, interactions felt friendly. I laughed authentically when sourcing for the female shadower, as opposed to nervously when I was sourcing for the male shadower. Questions were general and conversational: ‘Why do you want to work in this industry?’ or,

‘What was your exchange program like in Australia?’ It seemed that I had to say something erroneous before I lost the good favor of the interviewer. Through the female shadower I was relaxed to the point that I could enjoy the dialogue that took place in the interview and could build genuine rapport – a likely consequence of the affirmative responses and head nods I received.

Being interviewed as a man by other men was a novel experience. Interviews began in a collegial manner and indicative of a tone that I would expect in the locker room of a male sports team. With more senior interviewers, the tone often turned combative and I felt overwhelmingly defensive. Questions were pointed and challenging: ‘Why should I hire you?’ or ‘What value will you add to our team?’ I was climbing uphill from the start to prove that I was a worthy candidate for hire. I sat on the edge of my chair, sweating, as I did my best to respond to the litany of questions thrown at me. While the questions asked of the male shadower were more difficult, within them lay more opportunities to showcase expertise, so through the male shadower I had more openings to demonstrate industry knowledge than when I sourced for the female shadower.

To experience this setting as a first-person observer was a unique and robust means to gain insight into my research question. Without it, I never would have felt what it was like to be a fashion model, or treated like one of the guys, and to what extent such differences impact job interviews. This method allowed me to actively experience the perspective of another beyond what I could empathize on my own, and shed light on a series of other questions I might not have otherwise considered. Surely there are many practical applications in both research and the field that could entail experiencing the world through

the eyes of another (EC, third author).

Shadowing for a “Chat bot” Computer Program. The self-experiment in which GR (second author) participated involved dyadic social interactions involving a computer-human hybrid cyranoid. Rather than shadow for a human source, the researcher utilized Cleverbot, a web-based artificial intelligence program known as a “chat bot” designed to engage in text-based conversation (Carpenter 2014). Thus, the cyranoid in this study had the “body” of a human (GR) and the “mind” of a computer. Participants in the study were a mixture of university students and adults recruited from a large European city and were randomly assigned to one of two conditions: Know and Don’t Know. Both conditions involved participants engaging in 10-minute face-to-face dyadic conversations with GR (a 30-year-old female), and in both conditions GR merely shadowed the words Cleverbot provided while speaking none of her own thoughts. Participants were told they could speak to their interlocutor about whatever topic they liked. Participants in the Know condition were instructed before interlocution commenced that their interlocutor would be repeating words from a chat bot, while in the Don’t Know condition it was not until after the interactions were complete that participants learned their interlocutor had merely been repeating the words of a computer program.

To allow for spoken communication between the computer-human hybrid cyranoid and the participant, the services of another researcher were used. From a room adjacent to where GR engaged the participant, the second researcher listened to the conversation via a wireless microphone. When the participant spoke, this researcher would speed-type the participant’s speech into the Cleverbot program and then speak Cleverbot’s response to the input text aloud into a microphone connected to a covert radio relay that transferred audio to an inner-ear radio device worn by GR.

Reflection. I found out very quickly that shadowing for a computer program is not nearly as easy as I had envisioned it to be. The conversations deviated considerably from what one might consider ‘normal.’ The chat bot frequently generated completely irrelevant and downright nonsensical statements, leaving me in the position of having to say these things aloud while keeping a straight face. I was more conscious of my facial expressions during these interactions than I am normally,

and I tried to keep smiling no matter what the topic of conversation was or how strange conversations became. I had to suppress my laughter a number of times given the unpredictable responses the chat bot gave and the subsequent reactions by participants. At times I was frustrated with the inability to communicate my own thoughts and feelings to the participants, especially when the topic of conversation was something I was genuinely interested in. There were moments when it was particularly awkward, such as when the chat bot would discuss 'my' sexual orientation, or when male participants - knowing I was simply speaking what a computer program had generated in response to their statements - would flirt with me.

The experience has made me dramatically more aware of the mundane things that scaffold informal conversation. For instance, even in the condition where people knew I was simply repeating the words of a chat bot, participants continued to address me as if I was capable of responding in a human-like manner. I felt, in other words, that what persisted despite the abnormality of the situation was a strong desire on both my part and that of the participant to establish a human connection. In the face of repeated misunderstanding, there was a consistent urge to re-establish shared meaning. Furthermore, the experience has made me more aware of the role short term memory plays in buttressing understanding in conversations. The chat bot had an inability to remember the context of a conversation beyond several turns, thus conversations would fall into confusion whenever the participant would bring up something they or the chat bot had said at an earlier part of the conversation.

As a researcher, I feel that the experiment produced many important insights. I began to understand, to some extent, how

challenging it must be for someone who is limited in terms of language (e.g., by a communication disorder) to communicate with someone else who does not possess the capacity to understand them. I felt internally capable of establishing understanding with the participants but was incapable (by design, in this case) of actually doing so. Furthermore, the experience has made me more reflexive with regard to the role my outer identity plays in mediating communication, and has prompted me to want to look closer at the connection between attractiveness and perceived intelligence and its subsequent impact on psychological well-being. Drawing from the example of participants continuing to flirt with me despite the ridiculous words I was speaking: do a person's physical attributes lead another to look past their extreme inability to communicate thoughts in a coherent fashion? (GR, second author).

General Discussion

Benefits of Self-Experimentation. *Accessing Social Qualia.* Subjective self-reflections of the kind reported above illustrate the notion that via self-experimentation a researcher gains access to experience, which is itself a form of information that is not merely descriptive; it is experiential. Philosophers refer to this type of information as “qualia” (Jackson 1982), and it is often invoked in defense of the notion that conscious phenomena cannot be fully understood until they are known both in terms of their objective third-person description as well as through subjective first-person experience. Thomas Nagel’s (1974) “what is it like to be a bat?” thought experiment illustrates this point well: if all of the energies of every scientist in the world were directed toward studying the bat, it might be possible to arrive at a profoundly complete description of the species, yet the subjective experience of any one bat would remain utterly unknown. This paradox is present for human phenomena as well. For instance, in social psychology we now have many models and experimental validations describing social stereotyping and behavioral confirmation in relation to identity, but in what sense does reading this literature provide one with the experiential knowledge of being stereotyped? What is clear in the self-reflections provided in the current work is that each researcher had the ability to directly experience being stereotyped, whether due to age, gender, or physical appearance, and - importantly - the cyranoid method allowed the

researchers to systematically control the conditions that gave rise to certain forms of stereotyping. This is evidence for the utility of a researcher-as-subject social psychological experimental model in that such a model, in being a route to social qualia, could operate in conjunction with traditional experimental approaches to provide researchers with a more complete understanding of investigated phenomena - an understanding rooted in both description and experience.

Building Mental Models. Another benefit to self-experimentation is that it allows researchers to develop more precise mental models about the phenomena they study. By mental model we simply mean a person's internal representation of how various components of their environment (including social processes) operate in relation to each other (see Schumacher and Czerwinski 1992; Klimoski and Mohammed 1994). Mental models are acquired and developed via biological affordance, cultural learning, socialization, and lived experience (Johnson-Laird 1983), and it is to this last category that self-experiments can contribute, for they enable social researchers to directly experience the phenomena they wish to understand and thereby obtain a subjective and lived awareness of the properties involved. This is evident in the self-reflections shown in the present article, as each researcher noted how self-experimenting enhanced their internal representations of particular social phenomena that occur outside the experimental setting. Moreover, we suspect on the basis of the above-described self-reflections that there is a strong relationship between broadening and enhancing one's mental models with regard to social psychological phenomena and the creative process of developing new research questions (for further discussion on self-experimentation as a source of idea generation, see Cabanac 2004; Lubart and Mouchiroud 2004; Roberts 2004).

Enhanced Reflexivity. Doing self-experimentation is akin to being an ethnographer within one's own experiment, and the anthropological models of autoethnography and participant observation have much to offer a social psychological approach to self-experimentation. In particular, these research practices remind us of the maxim expressed in Habermas' (1968) classic work *Knowledge and Human Interests*: that we cannot escape our role in constructing knowledge and that the best we can do is be reflexive about our role in producing knowledge. To be reflexive means being able, to some extent, to step out from one's own role in both creating the research situation and constructing it. Especially in qualitative research it is common practice to extol the importance of reflexivity and critical thinking about one's own

role in research (Cornish, Zittoun, and Gillespie 2007; Denzin and Lincoln 2000). But, exactly how is one meant to be reflexive about one's own research? Building on the idea of position exchange (Gillespie and Martin 2014), that self-reflection can originate in exchanging roles or social positions, we propose that researchers moving between the social positions in their own experiment might be a viable route to becoming more reflexive about the given experiment.

In a sense, being the object of one's own experiment is a means of traversing the landscape of possible experiences and perspectives related to the different conditions within an experimental design. In the examples given in the current work, we see how a particular form of self-experimentation that made use of the cyranoid method enabled each researcher to become more reflexive with regard to the role of identity in shaping social experience as the various experimental conditions brought different aspects of identity to the fore. Given these results, it would be refreshing if it became common within experimental social psychology for researchers to report reflexive self-observations of experiments (even if only for pilot trials) together with standard observational third-person findings so that those who access this literature could become more fully attuned to the perspectival variety inherent in experimental scenarios.

Some Thoughts on How Social Psychologists Might Approach Self-Experimentation.

The distinction between laboratory and field experiments applies as much for researcher-as-subject designs as it does for traditional third-person designs. In the current work, Study 1 and Study 3 are both truer to the logic of traditional laboratory experiments in the sense that each researcher experienced systematic and repeated iterations of different experimental conditions in a controlled laboratory. By contrast, Study 2, having been conducted in an office environment, is more of a field experiment akin to Milgram's subway and waiting line studies, though it too involved the researcher experiencing repeated iterations of separate conditions involving participants with no explicit knowledge of the experimental manipulation. Thus, a researcher-as-subject model of self-experimentation seems viable for both laboratory and field studies.

Another issue concerns what, exactly, one should be observing in a self-experiment. Should one enter into a self-experiment with a predetermined set of dependent measures to systematically observe (e.g., one's psychological states)? Alternatively, should one adopt a

blank slate approach, beginning with a simple experimental design and prioritizing what to observe once in the midst of the experiment? Or should one approach self-experimentation not at all as a formal data-gathering exercise, but simply as a means of accessing experience? Our own perspective is that the answer to each of these questions is yes. None of the benefits of self-experimentation outlined above necessarily require any sort of formal data-gathering, yet there certainly can be utility in having specific dependent measures, particularly if a researcher's aim is to perform a combined analysis of first-person and third-person data. KC (Study 1) performed a content analysis using transcripts from the interactions to identify how their conversational engagement style compared between experimental conditions. Similarly, EC (Study 2) coded the quality of her responses to interviewers' questions to assess through which shadower (male vs. female) she more often produced ideal responses. GR (Study 3), meanwhile, recorded detailed notes following each experimental trial describing the subjective sensations she experienced. While it is true that third party observers could also have arrived at the data recorded by KC and EC (the data being overtly behavioral in nature), only the researchers themselves are in a position to make inferences based on the data that incorporate both first-person and third-person perspectives.

As a final note, researchers who do opt to incorporate formal data-gathering techniques as part of a self-experiment may want to consider the growing number of technologies designed specifically to capture first-person perspective data in social environments. These include the SubCam, a discreet head-mounted video camera that captures a social actor's visual field that has been used for first-person ethnographic analyses in occupational field studies as well as in experimental contexts (see Glăveanu and Lahlou 2012; Lahlou 2010, 2011). Other commercially available devices include lifelogging wearable technologies such as the BodyMedia FIT, designed to measure, among other things, sleep quality, galvanic skin response, and heart rate (for an overview of "quantified self" devices, see Lee 2013). Microsoft's SenseCam, a wearable camera that captures still images at adjustable intervals, has been used for first-person perspective analyses of routine activities related to, among other things, health and energy consumption (see Doherty et al. 2011; Gurrin et al. 2008) and has also been used to enhance the reflexive practices of teachers in classroom environments (Fleck and Fitzpatrick 2009). Each of these technologies captures a specific format of data (often capturing multiple formats simultaneously) and enables the researcher to go well beyond traditional self-reflection and introspection when conducting a self-experiment. The SubCam, for example, allows a subject (for our purposes, a researcher) to revisit their

subjective experiences and temporally reconstruct their psychological states “at the moment of action” (Lahlou 2011, p. 607). Such data, if collected methodically, can enable a variety of comparative analyses between conditions in a self-experiment. Quantified self devices have the advantage of being discreet, unobtrusive, and can be worn for long durations, thus enabling long-term self-experimentation as well as self-experimentation in remote field settings where the traditional tools of the lab may be unavailable.

Conclusion. There will always be an explanatory gap between the experience of the actor and the description of the observer (Farr 1996). The experience of an action and the observation of an action will necessarily feel different, with each yielding a distinctive form of knowledge. Recognizing this fracture as insurmountable entails opening social science up to a perspectival ontology, that is, the proposition that the social world comprises many potentially incommensurable perspectives (see Mead 1932). In the face of such plurality, what is the social scientist supposed to do? Rather than narrow the field of view to focus only on that which can be observed, we encourage experimenters to augment what is observed with what is experienced.

What we advocate is not so much an innovation as an appeal to acknowledge and legitimize what every good social experimenter does anyway, namely, walk through their own experiment, exploring the experiences that the protocol creates, and tweaking that protocol towards validity. Often this stage of research is relegated to the pilot phase and not written up. Yet, it is done; why? Because doing so is useful. Seeing an experiment from the point-of-view of a subject, among other things, helps us to understand the relationship between independent and dependent variables, to work out hypotheses, and to develop and organize new avenues of research. Self-experimenting helps us to understand ourselves better and reminds us of how important it is to confront the nature of experience created within an experiment directly. All we are suggesting is that this practice is necessary and valid, that it deserves its own conceptualization as a valuable tool on the mantle of acceptable, recognized modes of inquiry, and that it can be sought for its own ends.

CHAPTER 6 | Cyranoid Ethics

Those wishing to conduct social psychological research, as well as the institutional review boards (IRBs) that oversee such research, must perform due diligence in assessing whether a proposed study presents any ethical concerns. The principal ethical concern vis-à-vis the cyranoid method is the use of deception when using covert cyranoids (i.e., leading participants to believe they are speaking with an autonomous person when in reality they are in part communicating with an unseen person or machine whose identity remains hidden for at least some portion of the interaction). To gain IRB approval for the studies described in the previous chapters of this thesis, and to ensure participant wellbeing when carrying out the research, a range of protocols and experimental design features were operationalized (see Appendix D ethics material). This chapter outlines these various protocols and design features, details how participants reacted to the use of deception, and offers suggestions regarding cyranoid ethics that subsequent users of the method might consider in future research.

APA and BPS stance on the use of deception in psychological research

Both the American Psychological Association (APA) and the British Psychological Society (BPS) offer clear guidelines on the use of deception in psychological research in their respective codes of ethics (American Psychological Association, 2010; British Psychological Society, 2009). The APA compels researchers to only consider using deception when its use alone would enable a study to produce substantial scientific, educational, or applied knowledge, and forbids deception in cases where there is the reasonable expectation that it will cause severe physical pain or emotional distress. The APA further compels researchers to disclose the full extent of the deception to participants as early as feasible and no later than when the researcher finishes collecting data from them (e.g., during a debriefing session). Furthermore, researchers are obliged to grant participants the right to withdraw their data following debriefing. Similarly, the BPS asks that researchers avoid intentional deception unless it is exceptionally necessary to perform the research; if deception is used, additional safeguards must be considered to ensure participant wellbeing, and the participant must be informed of the deception as early as is feasible. The BPS makes the distinction between simply withholding details of a study from participants (e.g., the hypothesis or the exact nature of a stimulus) and deliberately lying to participants about the nature of the research, condemning the later. Both the APA and the BPS guidelines on the use of deception are

intended to safeguard not only the wellbeing of participants during and after their participation, but the trust participants place in the researcher, the researcher's organization, and the larger community of psychological researchers and research institutions as well.

One of core principles of the BPS's code of ethics and conduct is the concept of participant self-determination (British Psychological Society, 2009), which refers in part to each participant's right to determine whether or not a particular facet of the research they are participating in meets their own internal criteria for ethical appropriateness. The reaction each participant has upon being informed of deception, therefore, stands as sufficient without the need to compare the participant's response against some benchmark criteria (British Psychological Society, 2000), such as the reaction's comparability to other participant reactions or a predetermined scale set by the researcher. In other words, no matter how innocuous a certain deception may be in the mind of the researcher or in the minds of the majority of research participants, should one participant determine that their experience with the deception was unethical, then the deception was unethical. Likewise however, should participants see no problem whatsoever with a given deception upon being informed of its use during debriefing, then the researcher and the IRB are obliged to consider these views as valid and factor these views into any subsequent reapplication of the deception.

The nature of the deception used in this thesis

Given that each of the empirical chapters within this thesis involved the use of covert cyranoids, deception was employed to preserve the validity of experimental treatments, though not every participant was subjected to deception nor was every cyranoid covert in nature (some participants were specifically told prior to the experiment that their interlocutors would be cyranoids). In Chapter 2, "Replicating Milgram," participants in the treatment conditions (cyranoid conditions) of both studies were not told of the cyranic nature of their respective interlocutors prior to their interacting with them; withholding this information was required in order to both gauge the cyranic illusion as well as determine the extent to which the outer identity of an interlocutor mediated person perception and other behaviors. In Chapter 3, "Echoborgs: Cyranoids with computer program sources," participants in the treatment condition of the chapter's third study encountered a covert echoborg, and likewise were not made aware of the nature of their interlocutor prior to the experimental interaction. Similarly, in Chapter 4, "Using echoborgs to assess intersubjective effort in human-agent dialog," participants in the treatment condition encountered a covert echoborg and were thus

not made aware of the cyranic nature of their prior to the experimental interaction. This type of deception (i.e., withholding from a participant the true nature of a stimulus) is a common type of deception referred to by Greenberg and Folger (1988) as “deception to conceal the true purpose of the experiment,” which can be distinguished from other general types of deception such as “deception to conceal the true purpose of behavior” (i.e., when participants are misled about the true purpose of their actions), and “deception to conceal forthcoming experimental experiences” (i.e., when deception is used to keep subjects ignorant about certain events that will transpire over the duration of the experiment).

How deception was managed: Design and debriefing

In the undated transcription of his audio notes, Milgram (n. d.) conveys the following:

Thus it appears to me that the ethical issues in this experimental research center on the misrepresentation of the medium as being the source of his own ideas. For many subjects, it is possible to fully inform them of the technique that is being used, for we will want to know their reactions to the source and the medium when they are fully apprised of the situation. For other subjects, however, we will not want to inform them before the experiment of the cyranic arrangement. We will, however, give them as much information as possible while still allowing the experiments to be conducted. Specifically, they will be told that they are involved in an experiment, that there are certain features of the experiment that cannot be explained to them beforehand, but that they will be told afterward what the nature of the experiment was. All subjects will participate on the basis of voluntary consent (pp. 27-28).

Apart from this brief mention of the need to deceive participants in accordance with the particular research question being pursued and of the necessity of informed consent, no other archive documents written by either Milgram, his students, or his colleagues mention deception or the broader ethics of conducting cyranoid research. Thus, it was necessary to more or less consider these issues from scratch when first setting out to replicate Milgram’s early pilots and subsequently when expanding the cyranoid paradigm to include echoborgs.

When designing the replication experiments (Chapter 2), careful thought was given as to how to make the deception as benign as possible so as to reduce the chance participants would have a negative experience with the study or regret having been subjected to deception. The consent forms participants reviewed and signed prior to receiving instructions made it clear that there would be a full debriefing at the conclusion of the study during which the full nature of the study would be explained and that they would have the right to withdraw their data at that time should they so choose. The experimental interactions were designed to be informal so that participants would not feel under pressure to behave in accordance with a particular role or certain performance measures. No scripts were given, though the second study (adult-child cyranoid scenarios) did include a general set of guidelines for participants to follow in order to focus the conversation on intellectual matters. Both the research participants (who served as interactants) and the confederates (the sources) were informed that they should only speak about issues they felt comfortable speaking about and that they did not have to disclose any personal information. These instructions were provided so as to limit the possibility of conversations sliding toward taboo, vulgar, or personally uncomfortable topics. The final experimental procedures and instructions were settled on following repeated piloting of beta-version protocols so as to anticipate potential problem areas, such as the possibility of participant discomfort. When instructions were delivered to participants, they were expressly informed of their right to withdraw from the study for whatever reason even if the decision to withdraw interrupted the experimental interaction.

During the replication study's debriefing sessions, the researcher disclosed to participants the goals and motivations of the study, the full nature of the person they had interacted with, informed them of the identity of the source who had fed lines to the shadower (without disclosing any identifying personal information about the source), explained why the deception was necessary, and invited the participants to offer their thoughts on the deception. Allowing for an informal exchange between the researcher and the participants was intended to give participants the opportunity to fully understand the nature of the deception and to alleviate them of any concern of having acted inappropriately or insufficiently. Holmes (1976) refers to this type of debriefing as "desensitizing" as it is designed to help participants "deal with new information about themselves acquired as a consequence of the behaviors they exhibited during the experiment" (p. 868). According to Kidder (1981), one of the purposes of debriefing is to minimize any feeling amongst participants "that they might have been manipulated, made fools of, shown to be gullible, or revealed character weakness" (p.

404). It was decided that an effective way of gauging whether such feelings did occur was to grant participants the opportunity to voice whether they would object to someone such as themselves taking part in the same study under the same conditions. Thus, at the conclusion of the debrief session, a document was given to participants in which they were asked to indicate agreement or disagreement with each of the following statements (see Appendix D):

I (research participant) have been debriefed by the researchers as to the full nature of the full nature of the research study I have taken part in.

I understand that the nature of this study involved me speaking with an individual who was channelling the words of another person.

I would not object to another individual such as myself taking part in this study, should they consent to do so.

I consent to the researchers using all data collected from me.

I am aware of my right to anonymity with regards to this study and understand that the data I have given will be kept confidential.

IRB approval of the study was granted on the condition that participants' reactions upon being informed of the deception be closely monitored. Specifically, if at any time a participant voiced a concern regarding the ethics of the deception, or objected to another person being subject to the same deception, the experiment was to stop and no further trials were to be conducted until the deception protocols had been re-evaluated.

The ethical insights gained from the replication studies fed into the design of the covert echoborg scenarios featured in Chapter 3 (study 3) and Chapter 4. One of the anticipated challenges with the echoborg studies was that the interactions would be fundamentally different from those in the replication study, namely, the interactions would be a lot more awkward for participants. This is due to the fact that the dialog style of conversational agents

is generally quite incongruent from what most people would consider normal interlocution. Conversational agents often produce illogical and random statements, the content of which cannot be predicted. As such, the conversational agents that were employed (the chat bots Cleverbot, Rose, and Mitsuku) were vetted to determine whether participants would likely encounter anything vulgar or otherwise inappropriate. As with the replication study, the interactions with covert echoborgs were designed to be informal and relatively brief so as not to prolong any awkwardness, discomfort, or frustration felt by participants (covert echoborg interactions lasted 10-minutes, compared to the 20-minute interactions with the adult-child cyranoids in the replication studies). Protocols were thoroughly piloted with naïve volunteer participants prior to running the full experiments so as to vet the procedures and gauge likely reactions to a covert echoborg. Embarrassment and awkwardness were indeed experienced by these volunteer participants during piloting, but their reactions were overwhelmingly positive and they took a great interest in the echoborg concept when the deception was revealed by the researcher during pilot debriefing.

In the actual covert echoborg experiments, the debrief procedures were slightly different than those that had been employed for the replication studies. Participants were informed of the nature of the deception, namely, that the words the person whom had sat across from them spoke were determined by a conversational agent computer program and fed to shadower via a discrete radio transmitter. The reason for the use of deception was explained, and the researcher informally conversed with the participant about the motivations and goals behind the study. To conclude the debriefing session, the participant was verbally asked two questions by the researcher:

Now that you are aware of the full nature of the study, do you
wish to articulate any concerns you may have about the ethics
of such research?

Understanding the nature of the deception involved with this
study, would you object to someone such as yourself - for
instance, a friend, colleague, or family member - taking part in
the same study should they consent to do so?

As with the replication studies, IRB approval for the study was granted on the condition that

should any participant articulate an ethical concern regarding the study's protocols or object to another individual taking part in the study, the study would immediately halt and no additional trials would be run until the protocols had been re-evaluated.

Participants' reactions to the deception

Participants' reactions upon learning of the cyranic deception were overwhelmingly positive. A total of 134 participants from across each of the studies described in the empirical chapters of this thesis were subjected to deception (either encountering a covert cyranoid or echoborg), yet none raised so much as a slight objection to the deception upon its disclosure during debriefing. Most participants expressed genuine interest in the method, and many stayed beyond their time slot to further discuss details of the study and the objectives of the research. Among those who encountered an echoborg, many participants likened their experience to that depicted in popular science fiction films that they had seen (a frequent reference was the recent Academy Award-winning film *Her*; Jonze, Ellison, & Landay, 2013). To give a sense of participants' feelings toward the most extreme use of cyranic deception used in this thesis, Table 6.1 documents feelings regarding the ethical nature of the research expressed during debriefing by each of the participants that encountered a covert echoborg in the empirical study documented in Chapter 4.

Ethics considerations for future cyranoid research

Taking stock of the entirety of first-hand experiences with the cyranoid method gleaned from the various studies within this thesis, a number of suggestions with regard to research ethics can be offered to those considering future cyranoid research.

Consider whether deception is at all necessary to the research question. As demonstrated in Chapter 3 of this thesis, not every research question that necessitates use of the cyranoid method likewise necessitates withholding from participant interactants the cyranic nature of an interlocutor. Indeed, such deception is generally only necessary if it the researcher's goal to investigate a cyranic context in which the interactant is in "default mode" (i.e., wherein they naturally assume that persons they encounter are communicating self-authored, spontaneous thoughts).

Consider the behaviors likely to be exhibited by the participant and whether participants might regret exhibiting such behaviors upon being debriefed. As is usually

Table 6.1

Participants' Thoughts on the Ethical Nature of the Study (Chapter 4: "covert echoborg" condition)

Trial	Gender (age)	Do you have any concerns regarding the ethics of this research? ^A	Would you object to someone else like yourself taking part in this research? ^B
1	Female (27)	No I think it has to be done this way. For me it's okay to know now what's behind it.	No not at all. I would encourage it.
2	Male (21)	I think it was fun.	No no. Obviously not.
3	Female (21)	No. I think it was necessary for you to do that, otherwise I would have acted in a very different way. So I think it's okay.	No I wouldn't.
4	Female (22)	The study felt ethical. There was no coercion, no forcing. So, ya, I think it felt ethical. It was a good experiment, ya.	No. Not at all.
5	Female (19)	No. I was uncomfortable doing it but now I'm fine.	No.
6	Male (24)	I was totally comfortable.	Not at all. Not at all.
7	Female (23)	No, because I think it's interesting to know how people react in those situations. So no, it's fine.	No.
8	Female (23)	Ya it's okay. It was quite interesting actually.	No.
9	Female (20)	Ya, I mean this might be the way artificial intelligence is heading. I'm not ethically opposed to it. In the social sciences you gotta do this or you'll get a totally different response.	No. I'd feel bad because that was a monstrously uncomfortable 10-minutes. No, but I think it'd be kind of fun to laugh about it afterwards.
10	Female (22)	No it was fine.	No I wouldn't because I think they'd be aware this is an experiment so I think withholding this kind of information is fine.
11	Female (25)	No not really.	No.
12	Female (29)	Ya I was happy, it was fine. No problems.	No, it'd be fine.
13	Female (20)	I wouldn't say anything was ethically wrong with it, I just felt a bit uncomfortable <laughs> in the conversation, but like, that's the conversation, like... I feel fine about it	No.
14	Male (23)	Ya, I agree that my answers would be different if I knew. Ya, I completely	No.

		agree and I don't see any problems.	
15	Female (23)	Ya I feel it was ethical. I was just following the procedures. I didn't feel anything strange or unethical about it.	No.
16	Female (20)	No it was fine.	No.
17	Female (36)	No it was fine.	No, no, it's fine.
18	Male (30)	No I didn't feel uncomfortable at any point. I'm happy with the study.	No.
19	Female (19)	No it was completely fine.	No.
20	Female (28)	No, I think it was great.	No, of course.
21	Male (26)	No not at all. If you told me the results would not be the same.	No no.
22	Male (20)	I'm pretty comfortable. I can see how some people might find it frustrating but I wouldn't think it'd be a significant concern.	I wouldn't object to that.
23	Female (18)	No. I'm okay with it.	No, I would just wait until they found out to see what they said <laughs>
24	Male (23)	No.	No.
25	Female (21)	Ya I'm okay with the procedures. I think that's how you get, like how people react in real life.	No. I wouldn't even tell them about this <laughs>
26	Female (19)	No. Everything's fine.	No.
27	Male (21)	I thought it was fine, there wasn't anything unethical about it.	No, I don't see any reason why I should object to that.
28	Female (25)	No.	No.
29	Male (21)	Well if you had told me you wouldn't have been able to conduct the study.	No.

^A Full question: "Now that you are aware of the full nature of the study, do you wish to articulate any concerns you may have about the ethics of such research?"

^B Full question: "Understanding the nature of the deception involved with this study, would you object to someone such as yourself - for instance, a friend, colleague, or family member - taking part in the same study should they consent to do so?"

the case in person perception social research, experimenters look for explicit and implicit evidence as to whether a research participant has confirmed a particular stereotype or otherwise exhibited behavior or volunteered judgements influenced by some overt characteristic (e.g., physical identity) of a stimulus person. This is typical of research involving phenomena such as persuasion and attitude change, for example. Upon learning of

the intentions of a study during debriefing, therefore, there exists the possibility that participants will come to evaluate the behavior they exhibited or the responses the indicated in a negative light being that they might have confirmed some socially taboo stereotype. Given that a goal of ethical research design is to mitigate the possibility that participants will leave a study feeling self-critical about their performance (Kidder, 1981), serious thought should be considered prior to using deception to conduct cyranoid research regarding whether or not the designed interaction scenario combined with the deception is likely to elicit behaviors that the participant will likely regret.

It is easy to imagine cyranic scenarios, for example, in which a participant interlocutor's belief that they are interacting with one type of person leads them to say or do something they wouldn't otherwise say or do to a different type of individual for reasons that have to do with social etiquette, cultural standards, in-group/out-group norms, and so on. The goal of cyranoid research should not be to elicit these types of behaviors full-bore without first considering what impact this might have on participants following revelation of the deception during debriefing. To not consider this possible impact would grossly violate the rights of the participant and undermine their trust in social psychological researchers and research institutions. This isn't to say that one cannot use the cyranoid method to investigate behavioral and perceptual social biases – the empirical studies in this thesis did just that. On the contrary, there are ways to design cyranoid experiments that gauge people's social biases without necessarily causing the participant to perform behavior that they end up regretting. The interaction scenarios involving covert cyranoids employed in this thesis, for example, featured interactions designed as cordial, informal conversations, and participants were encouraged not to say anything their interlocutor might find vulgar or inappropriate. Participants were not instructed to compete with their interlocutors, nor were they encouraged by the researcher to converse on taboo subject matter (e.g., participants in the child-source/adult-shadower cyranoid condition from the study in Chapter 2 were not encouraged to talk about "the problem with kids these days").

Pick sources wisely, and instruct them on how they should *not* behave. As the cyranoid method can be used to assess the behavior and perception of participants *as sources*, thought must be given to the possibility that a source might abuse their status as the determiner of the words spoken by the shadower by compelling the shadower to utter things in the presence of interactants that are inappropriate, cruel, or otherwise ethically problematic. This was not an

issue in the cyranoid experiments described in this thesis as sources were confederates who understood not to take advantage of the shadower in such a way, nor was this an issue in the echoborg extension of the cyranoid method given that the chat bots that were selected were not programmed to be abusive (the chat bots weren't even aware they were controlling a shadower, for that matter). Nonetheless, it is easy to imagine how an inappropriately behaving source can cause negative experiences for the other people involved in a cyranic interaction. This is not to say that the source's behavior be overly restricted. On the contrary, it is quite an interesting research question to consider how sources, from their inherent positions of anonymity, might behave in a manner incongruent with how they would behave in person depending on the nature of the cyranic scenario and the identities of the shadower and interactant. Nonetheless, any experiment involving participants being assigned as sources should have thorough protocols for making sure they understand what types of behaviors are unacceptable (e.g., instructing their shadower to utter derogatory remarks).

Document participant reactions to deception. Whenever the cyranoid method is deployed with deception, it is highly recommended that participant reactions to the deception during debriefing be well-documented, analyzed, and reported where possible (e.g., Table 6.1). This practice can serve two general functions. First, it can help researchers to evaluate their use of deception retrospectively and factor this evaluation into new research designs. Second, the transparency that such a practice would offer to members of the public, other researchers, journalists, institutional review boards, and other agencies would likely empower researchers hoping to conduct studies using the cyranoid method to see their research designs gain approval and funding.

Enhancing participation with the method. One thing that went unexplored in the current thesis is the possibility of allowing research participants to participate as a cyranoid during or after debriefing, particularly if the study involved deceiving the participant. Many of the participants that took part in this thesis' various studies were very inquisitive during debriefing as to how the cyranoid method worked in practice, particularly in the case of the echoborg experiments. Given the interest in and enthusiasm for the method that was expressed by the majority of participants, allowing all participants to have hands-on experience as a cyranoid might go a long way toward attenuating any unexpressed negative attitudes corresponding to their being deceived, or simply allow them to entertain their curiosity. Indeed, it is often the case that informed consent forms indicate that the participant

is likely to receive an educational benefit from their participation (i.e., the opportunity to learn something about science and, possibly, about themselves), and debriefing is the space in which the researcher can communicate the value in what participant has experienced (Tesch, 1977). Allowing participants to experience being part of a cyranoid could therefore help fulfil this educational service objective.

Field studies and other research variants. The mundane realism of a laboratory experiment plays a significant role in enhancing the ecological validity of the experiment's findings (i.e., the extent to which the laboratory results reflect what actually occurs in people's ordinary social environments), which is why the cyranoid method poses substantial potential as a means of experimentation. The importance of assessing ecological validity is a reason why social researchers often bring an experiment or intervention of some sort out of the laboratory and into the "real world" in the form of field experiments or field studies. Apart from the field experiment described in the first-person reflection by EC in Chapter 5, all of the experiments contained in this thesis made use of the cyranoid method within controlled laboratory settings. The cyranoid method has been deployed in small field studies conducted by Robb Mitchell and colleagues, such as in art installation and workplace settings (Mitchell, 2009) and classroom environments (Mitchell, 2010; Raudaskoski & Mitchell, 2013). Nonetheless, many facets of the cyranoid method (including the echoborg) remain to be investigated outside of the research laboratory. Field research of any kind, however, poses many unique ethical challenges, and cyranoid field research bears no exception.

One of the chief concerns with field research is the form of consent participants are able to give prior to being studied. As field research is often interested in how people behave in the real world, granting informed consent might impact participants' subsequent behaviors to the point of compromising validity. For this reason, according to Reis and Gosling (2010), ethics committees are often flexible on the scope of informed consent in field studies, as "consent can typically be bypassed in studies that are solely observational and that involve anonymous, public behavior" (p. 90); they also describe that when a study involves an intervention (e.g., a manipulation of some sort) prior to or in the absence of informed consent, "researchers must take more than the usual amount of caution to ensure that participants will not be harmed, distressed, annoyed, or embarrassed" (p. 90). Furthermore, Bouchard (1976) suggests that, in the field, the effects of independent variables and the variation in these effects are likely to be more pronounced than they would be in the laboratory, meaning that it can be more difficult

to foresee prior to conducting a field study all the ways in which a stimulus might impact participants, including their seeing the administration of a stimulus as unethical.

If participants are aware of their encountering an overt cyranoid in the field, which has been the case in past usages of the cyranoid method in participatory interactive research (e.g., Raudaskoski & Mitchell, 2013), these ethical challenges are significantly less pronounced as participants are fully aware of the nature of the cyranic encounter they find themselves in (though precautions should still be made by the researcher to ensure that sources do not abuse their control of shadowers or abuse interactants from their position of anonymity). Deploying *covert* cyranoids in the field in such a way that they engage with unsuspecting individuals, on the other hand, is an entirely different matter, particularly if these individuals have not granted prior consent. It would run counter to the contemporary standards of ethics in social psychological research to deploy a covert cyranoid in such a way unless it could be assured to a high degree that the cyranic encounters would be more or less innocuous, limited in duration, and performed in a social environment in which it was easy for the individual who encountered the cyranoid to relocate themselves. As an example, one of the studies that was considered for the current thesis was a “barista echoborg” field research scenario in which a chat bot programmed to respond to orders for coffee and make small talk was paired with a human shadower capable of physically operating café machinery (e.g., an espresso machine). This echoborg would be deployed in a public setting (e.g., a university quad) and participant reactions to the echoborg would be observed. Such a scenario was considered to be ethically acceptable given the likely harmless nature of the cyranic encounters and their limited duration.

CONCLUSION

The overarching aim of this thesis has been to lay the groundwork for future research involving the cyranoid method in various domains connected to social psychology. To achieve this primary aim, various sub-aims were pursued across the various chapters. In the concluding portions of the thesis that follow, each of these aims will be addressed. Discussions on limitations and future research areas are provided.

Aim: Present the first comprehensive introduction to the cyranoid method and situate it historically

The first sub-aim of this thesis sought to retrace Milgram's steps with cyranoids and place the method within social psychology's broader historical context so as to arrive at an explanation for the method's dormancy and posit a way forward for the method as a legitimate contemporary approach to experimentation. Archival materials show that Milgram devised the cyranoid method in the fall of 1977 in collaboration with graduate students from his mass media course at the CUNY Graduate Center in Manhattan (Milgram, 1977b). Initial applications of the method involved Milgram and his students exploring various cyranic permutations (e.g., a single source, multiple sources, covert and overt cyranoids, and so on) in largely exploratory small scale pilots; their fascination with the technique was that it seemed a highly innovative way of pairing the mind (or, words) of one or more individuals with the body of another (the speech shadower) in a way that created the illusion that the shadower was spontaneously speaking their own self-authored thoughts (Milgram, 1977a). This approach to generating new research methodologies was typical of Milgram as he often relied on the creative and collaborative dynamics of his classroom environment when choosing what social psychological phenomena to pursue (Takooshian, 2000).

Shortly after he devised the cyranoid technique, Milgram unsuccessfully applied for a large grant from the United States' National Science Foundation. The grant application failed principally due to Milgram's inability to articulate a clear theoretical focus and the extremely unconvincing research designs he proposed, which, among other things, included him suggesting that the cyranoid method be used in mock hostage negotiation simulations that would be filmed and shown to research participants – a clearly more theatrical than scholarly pursuit (Blass, 2004; Milgram, 1979). One of Milgram's critical errors was his inability to sell the method as an interactive methodology that permitted face-to-face encounters between

research participants and a person stimulus, a main affordance of the method that differentiates it from more static approaches to exploring the relationship between social perception and inner vs. outer identity. Despite failing to secure the NHS grant, Milgram persisted with the method. He eventually secured a modest \$5,000 grant from CUNY to pursue cyranoid research from 1983 to 1984, and with this grant conducted his most formal cyranic pilots, two of which are detailed in a speech he pre-recorded for an APA conference symposium in the summer of 1984 (Milgram, 2010a). In December 1984, several months after his APA presentation, Milgram died of a heart attack, and with his death came a period of dormancy for the cyranoid method that has to this day persisted within social psychology.

Beyond Milgram's grant application shortcomings, this thesis has accounted for the dormancy of the cyranoid method by considering how Milgram's style of social psychological research fell out of fashion during the very era in which he attempted to market the method. Milgram's research approach stemmed from the situationist perspective he adopted from his intellectual forbearers – social psychologists such as Sherif, Asch, and Garfinkel (Martin, 2015). His interest was in how the power of social situations shaped people's behavior more so than the influence of internal traits and personality dimensions. This interest led Milgram to adopt an experimental style that focused on observing social behaviors within installation-like settings, whereby confederates, props, and staging created a dynamic situation that the research participant would have to cope with and respond to. Cyranoids were emblematic of this approach; age-incongruent cyranoids in particular created a situation in which participants had to interact with a stimulus person that radically diverged from social norms. By the 1980s, however, this approach to social psychological experimentation had become largely eclipsed by a predominantly cognitivist approach to experimental social psychology that favored static and generally more reductionist experimental methods, privileged internal validity over mundane realism, and took more interest in constructing predictive models of human behavior, emotion, and attitude (Adair, 1991). Moreover, at the time of Milgram's cyranoid research, social psychologists were largely turning away from a dependence on deception of the kind Milgram's methods were often known for. Thus, right as Milgram was attempting to legitimize cyranoids as a research tool, his field was undergoing a dramatic paradigm shift, the consequences of which did not bode well for Milgram-esque research broadly and the cyranoid method specifically.

This thesis has highlighted a contemporary methodology that is analogous in many ways to

the cyranoid method and that can serve as an analogue for modern-era cyranoid research: immersive virtual environment technology (IVET). IVET proponents (e.g., Blascovich et al., 2002) argue that the method offers an opportunity to conduct social psychological experiments high in both mundane realism and internal control, given that participants in an IVET experiment interact within dynamic social contexts and can have their outer identities (i.e., avatars) systematically manipulated. It has been argued that this same logic applies to the cyranoid method, only with cyranoids researchers can preserve the ability to have research participants encounter real person stimuli as opposed to virtual stimuli.

Aim: Validate the cyranic illusion and explore its robustness

In each study that investigated the cyranic illusion, participants largely, if not entirely, failed to detect that the person they physically encountered (the shadower) was not speaking self-authored thoughts. In the case of covert echoborgs, however, many participants did suspect that their interlocutor was deviating from their normal way of behaving (though, when pressed, these participants failed to go so far as to articulate suspicion that their interlocutor was being told what to say by a third-party). This of course is largely a consequence of the present limitations of chat bots more so than a limitation of the echoborg method in and of itself being that conversational agents cannot, at present, fluidly converse in a manner typical of mundane human-human dialog. Should a particular conversational agent someday become so advanced as to be indistinguishable from a human interlocutor, then there is every reason to suspect that interactants who encounter covert echoborgs will succumb to the cyranic illusion every bit as fully as those who succumb to the cyranic illusion in contexts involving congruent source-shadower cyranoids.

It is not at all surprising that people succumb to the cyranic illusion. After all, most people for good reason do not go about their daily lives suspecting that those they encounter are receiving words from a remote third party. It would be somewhat facile, therefore, to pursue research questions that purely concern under what conditions people are more or less likely to notice an interlocutor that is a speech shadower, as there is good reason to suspect that most people would fail to notice such a situation even when confronted with a shadower whose identity was markedly incongruent from its source. The cyranic illusion does, however, predicate much of the utility of the cyranoid method, particularly in circumstances involving interactants engaging with a covert cyranoid, for it is the assumption that an interlocutor is acting autonomously that enables the method to be useful when researching how the

disposition of a source is perceived and interacted with on account of the identity of the shadower they are paired with. Therefore, exploring the robustness of the cyranic illusion, particularly under extreme conditions (e.g., a covert echoborg), was necessary for the purposes of the groundwork laid out in this thesis.

Why do people succumb to the cyranic illusion even when instructed to expect the possibility that an interlocutor is not communicating their own words? This question was explored in Chapter 3, Study 2. People were confronted with either a text terminal or an echoborg and asked to determine over the course of a 10-minute interaction whether or not they believed their interlocutor was truly communicating words authored by a chat bot or simply imitating one (but speaking self-authored thoughts nonetheless). As opposed to those who engaged with a text terminal, those who engaged an echoborg were far more likely to conclude that their interlocutor was simply imitating a chat bot. The concept of perceptual salience helps account this finding. Perceptual salience suggests that people will deem causal what is most salient to them in the absence of other salient explanations (Taylor & Fiske, 1975). People confronted with an echoborg (or a cyranoid) *see* in front of them a human producing speech and performing nonverbal behavior that matches that speech, leading them to attribute verbal agency to the shadower when this agency in reality rests with a remote source.

Aim: Demonstrate the cyranoid method as a means of conducting human-agent interaction research

This thesis extended the cyranoid method well-beyond the scope Milgram envisioned by introducing the echoborg, a special type of cyranoid composed of a human speech shadower paired with an artificial conversational agent source. As discussed in Chapter 3, humans have throughout history only ever interacted with machine intelligence through mechanical/artificial interfaces (e.g., screens, robotic devices, audio response systems, and so on). The echoborg enables human-computer interaction through a totally human interface whereby interactants engage with machine intelligence under the same social psychological conditions that characterize the primordial human interface: human bodies interacting in person and face-to-face. As demonstrated in Chapter 3 and Chapter 4, with echoborgs researchers can explore how the human body, as the delivery mechanism of communication, fundamentally alters perception of, and interaction with, machine intelligence.

The echoborg completes what was previously an incomplete matrix of interactant/stimuli-

types in the field of android science, the interdisciplinary domain that seeks to understand the social and cognitive phenomena emergent from human interactions with humanlike interlocutors (MacDorman and Ishiguro, 2006b). Prior to the introduction of the echoborg, android science involved interactions between humans (mind: human; body: human), autonomous androids (mind: machine; body: machine), and tele-operated androids (mind: human; body: machine). The echoborg inverts the composition of the tele-operated android, achieving an entity composed of a machine mind and a human body (insofar as “mind” is operationalized as the component of an entity that produces words and interprets the words of others).

The echoborg contribution to android science is important for several reasons. First and foremost, there are currently many technological constraints to achieving a complete machine imitation of a human being (i.e., creating a human being that can pass a *Total Turing Test*, as defined by Harnad, 1991). Chapter 3 discusses two general categories of constraints: (1) those that are born of the difficulty in arriving at conversational agent technology that can participate in mixed-initiative casual conversation typical of mundane human-human interaction, and (2) those that are born of the difficulty in synthetically recreating the motor capacities and physical attributes of the human body. Whereas the tele-operated android constitutes a trade-off whereby the former category of constraints is nullified on account of a human operator that controls the android’s speech and motor behavior (Nishio, Ishiguro, & Hagita, 2007a), the echoborg nullifies the latter category of constraints by involving an actual human body rather than an imperfect machine imitation of one. The echoborg, therefore, to some degree “leapfrogs” the contemporary inability to create a machine with the appearance and motor skills human and provides for an experimental person stimulus that is an approximation of the ideal scenario: a machine with a body indistinguishable from that of a real human. Another important affordance the echoborg lends to android science is that it allows for a unique approach to understanding the “uncanny valley,” the hypothesis first suggested by Mori (1970) that motivates much of the empirical work within the field. Until now, social psychological experiments that have explored the uncanny valley have focused almost exclusively on the appearance of artificial agent and/or the motor behavior the agent. The echoborg, however, provides for a perfectly human stimulus in terms of appearance and motor behavior, allowing researchers to study whether an uncanny valley emerges in relation to the humanlike or non-humanlike interlocution capabilities of an artificial conversational agent, and whether the uncanny effects of an artificial agent are moderated by the interface

they communicate with (e.g., human body vs. mechanical body). In Chapter 3, Study 3, various self-report measures typically associated with uncanny valley research were explored, and it was found that interactants who engaged with a covert echoborg felt significantly *less* comfortable speaking to their interlocutor than interactants who engaged the same conversational agent but through a text interface, suggesting that an agent that violates the expectations typically associated with human-human interlocution can produce uncanny effects *despite* being perfectly humanlike in terms of appearance and motor behavior.

Aim: Show how echoborgs can be used to gauge intersubjective phenomena emergent in human-agent interaction

Whereas Chapter 3 was very much a proof-of-concept concerning the echoborg extension of the cyranoid method, Chapter 4 utilized the echoborg to deeply explore the intersubjective phenomena emergent in human-agent interaction (namely, efforts toward establishing common ground through spoken language). One of the historical drives of android science (as well as many other domains within social robotics and intelligent virtual agent development) has been the assumption that as an agent's means of interfacing becomes more humanlike, so too will the behaviors and perceptions the agent elicits from its human interactants (Krach et al., 2008; MacDorman, 2006). Chapter 4 demonstrated that an important factor modulates the effect of the interface: the role of belief. Results showed that an agent-interlocutor that has a human body (i.e., a truly human means of interfacing) will not elicit the same degree of intersubjective behaviors from an interactant if the interactant knows that their interlocutor's communication is agent-determined (as opposed to human-determined). The echoborg method was demonstrated as a unique means of teasing out the influence of belief because it is the only method whereby a person can interact face-to-face and in person with a human-bodied interlocutor and believe they are speaking to an autonomously-communicating person yet in reality be engaging the words of a conversational agent. As it turns out, both a human interface *and* the belief that an interlocutor is an autonomously communicating person are required in order for an interactant to exert benchmark intersubjective effort, which can be operationalized as the extent to which a person attempts to repair misunderstandings as they arise with an interlocutor. This finding is important for our understanding of human-agent interaction, and may suggest that no matter how humanlike interfaces become, humans will only attempt to attain common ground with an agent at human-human benchmark levels if they assume the agent is actually a human. This finding also presents an important methodological contribution to the ongoing effort to establish means of benchmarking

human-agent interaction using frameworks tied to intersubjectivity theory (see Cassell & Tartaro, 2007).

This thesis demonstrated that the echoborg, and the cyranoid method more generally, can be important tools in human-agent interaction research that explores fundamentally social psychological processes such as conversational repair and perspective-taking. The echoborg gives researchers the ability to present research participants with person stimuli that elicit a stronger reaction than traditional stimuli – which in the domain of human-agent interaction are often text interfaces or virtual representations of humans or humanlike characters presented on computer screens or within immersive virtual environments. Analogous to the echoborg’s relationship with the tele-operated android paradigm, the echoborg is similarly an inversion the “Wizard of Oz” technique commonly deployed in virtual agent research whereby a research participant is told that an interlocutor is an autonomous agent when in reality the agent is being remotely controlled by a human operator (Dahlbäck, Jönsson, & Ahrenberg, 1993). By having a human confederate be controlled by an agent, researchers can test how the human body as an interface mediates dispositional elements of the agent. Furthermore, the echoborg can be used to bypass the need to engineer a virtual interface to be paired with a conversational agent; if a developer’s goal is to test how a particular piece of conversational software performs when in dialog with a human interactant and desires for the interactant to encounter a human body so as to induce the social psychological states associated with human-human interaction, the echoborg method eliminates the need to develop a virtual agent interface (which is time consuming, expensive, and, and in any case is an imperfect approximation of an actual human interface).

In demonstrating the echoborg’s affordances for exploring human-agent intersubjectivity, this thesis also contributes to quantitative intersubjectivity methodology by demonstrating how dependent variables operationalized via Conversation Analysis techniques (namely, other-initiations of repair and repairs proper) can be examined within a standard experimental framework. Conversation Analysis researchers rarely use quantitative methods such as those applied in Chapter 4 (Schegloff, 1993). This thesis, however, showed how multilevel logistic regression can be used to determine the likelihood that an utterance articulated by one interlocutor within a dyad *within an experimental condition* will be followed by an other-initiation of repair, and this likelihood can be compared across experimental conditions so as to determine the effect certain experimental manipulations have on the relative incidence of

repair initiations. It is hoped that this process of statistically analyzing repair phenomenon can provide a model for future research aimed at assessing changes in intersubjectivity induced by treatments in social psychological experiments. Furthermore, it is hoped that the combining of the echoborg method with statistical operationalizations of intersubjective effort (e.g., repair-after-next-turn dynamics) can serve as a template for methodologies that aim to benchmark the interactive dynamics between humans and artificial agents (e.g., Cassell & Tartaro, 2007).

Aim: Argue that the cyranoid method can be used as a first-person means of conducting social psychological research

A formal revisiting of the cyranoid method would be incomplete without exploring the experiences of the individuals who actually compose a cyranoid or echoborg: the source and shadower. One could design an experimental study wherein research participants were randomly assigned to these roles and evaluated much like any other experimental protocol. However, in the spirit of Milgram's penchant for first-person, experiential research (e.g., Milgram, 2010c, 2010d), and given the unique affordances of the method for enabling people to experience social interactions given a radically altered outer or inner identity, the approach settled on in this thesis was to advocate for the cyranoid method as a means of researcher-as-subject self-experimentation. First-person, researcher-as-subject research has largely been resigned to social science domains adjacent to, but not necessarily located within, experimental social psychology (see Danziger, 1990). As discussed herein, this relegation has occurred for largely historical reasons that stem both from a mid-20th century institutional convergence on an interest in the distribution of psychological phenomena manifest in populations (as opposed to the fundamental nature of subjective experience) as well as long-held disciplinary anxieties about the status of social psychology's methods relative to those of other scientific domains (Danziger, 1990; Farr, 1978, 1983). Though these anxieties are in many respects well-founded, the possibility of first-person experimental methods becomes less taboo when the traditionally narrow goals of experimentation are augmented so as to incorporate aims beyond merely ascertaining the averages and variances of dependent measures. As argued herein, gaining subjective knowledge of a particular social phenomenon (e.g., stereotyping on the basis of outer identity) can be a legitimate goal of experimentation in and of itself in a manner that does not conflict with knowledge claims made via traditional third-person experimentation.

This thesis suggested three potential benefits for researcher-as-subject research that makes use of the cyranoid method: (1) access to “social qualia” (i.e., the subjective experience associated with particular social phenomena, such as stereotyping), (2) an improved ability to build mental models corresponding to social phenomena, and (3) an enhanced ability to be reflexive about social experiments and the role of identity in shaping the experience of an experiment. The journal article developed for Chapter 5 was written in a spirit of hoping to initiate what could someday be a larger discussion about these and other potential benefits of researcher-as-subject approaches to experimental social psychology, and the cyranoid method was offered as merely one of many possible means of conducting such research. The cyranoid method’s particular strengths as a technique for conducting first-person experimentation connects with Green’s (1983) discussion of the method as a technique for enhancing self-knowledge through physically transforming the material outer presentation of the self. Popper discussed this in philosophical terms when he argued that self-observation in contexts of transformed material manifestations of the self can lead to enhanced self-knowledge (Popper & Eccles, 1977).

Though Chapter 5 merely offers self-reflections from social psychologists who have participated as either a source or a shadower in cyranoid studies as evidence, there are certainly other forms of data collection and data presentation suited for first-person, research-as-subject experiments. Digital first-person ethnographic technologies, such as subcams (Lahlou, 2011), quantified-self devices (see Lee, 2013), and wearable still-image cameras (e.g., the Microsoft SenseCam; see Doherty et al., 2011) can be incorporated into future research-as-subject social psychological self-experimentation to give the researcher a broader means systematizing their observations.

Aim: Overview the ethics issues involved with cyranoid research and outline suggestions for avoiding ethics pitfalls

The final sub-aim of this thesis involved putting forth an initial ethics framework for the cyranoid method. The main ethical point of concern with regard to the use of covert cyranoids (or echoborgs) is that deception is required in order to preserve the interactant’s belief that they are encountering an autonomously communicating person. Milgram only very briefly grappled with the potential problems associated with this type of deception (see Milgram, n. d.) and did not systematically detail participants’ thoughts on its use (though it is fair to say that given the innocuous nature of Milgram’s cyranoid pilots, it is unlikely that he

encountered any participants who felt negatively toward having been deceived). It is important to keep in mind that many interactive social psychological experiments involve deception of some kind, particularly those that involve some form of mediated interpersonal communication. The Wizard of Oz technique common to many human-agent interaction studies, as mentioned before, involves leading a research participant to believe they are encountering an artificial agent when in reality the agent is controlled to a certain degree by the researcher or some other party to the experiment. The particular concern with covert cyranoids, however, is that the presence of another physical human (the shadower) in the room as a person stimulus can evoke much stronger and more visceral reactions to the experimental scenario than more technologically mediated, static, or distal person stimuli. As demonstrated in Chapter 3, Study 3, for example, interactants feel far less comfortable interacting with a covert echoborg than they do interacting with a covert text bot, even when the particular conversational agent generating the words received by the interactant is the same in both contexts.

One can easily imagine situations where research participants would feel regret at having said certain things or acted in certain ways toward a shadower upon learning of the identity of the source during debriefing. For example, it would be fairly straightforward to design a cyranoid study aimed at gauging certain forms of prejudice wherein interactants felt comfortable revealing hostile attitudes toward a particular demographic of people given the identity of the shadower they were asked to converse with. Participants should not feel any undue embarrassment at having been deceived or leave the research laboratory feeling as though they were tricked into displaying regrettable behavior. Therefore, it is vital to always proceed with caution when designing experimental scenarios involving covert cyranoids.

Chapter 6 outlines a series of suggestions for proceeding with deception in a cyranoid study. These include (1) determining the extent to which deception is even necessary for a given cyranoid experiment, (2) avoiding constructing scenarios likely to elicit regrettable participant behavior, (3) carefully picking sources and instructing them on how *not* to behave when paired with a shadower, (4) documenting participant reactions to deception during debriefing (e.g., Table 6.1), (5) giving participants who were subject to deception the opportunity to function as a shadower or source so as to enhance their participation in the research, and (6) avoiding field studies that unfairly take advantage of unsuspecting participants. Adopting each of these protocols when conducting cyranoid research will help

ensure positive participant experiences and help to maintain the method's reputation among the wider community of experimental social psychologists.

Limitations of the current work

1. Sample Sizes. Sample sizes for the various experiments conducted for this thesis were small. There exists the possibility, therefore, that some of the non-significant findings are Type II errors: failures to detect the effects of a manipulation when effects were actually present. Calculating the statistically appropriate sample size for an experiment is directly tied to the effect size one hypothesizes finding for a given manipulation as well as anticipating the standard deviation of dependent measures (Field, 2009). Since the experiments described in this thesis were largely exploratory, no assumptions regarding effect sizes or dependent measure variance were taken into consideration prior to designing and running the experiments, and the sample sizes that were achieved were deemed sufficient. On one hand, however, it is quite common for studies that introduce a novel methodology, particularly within fields that touch upon human-agent interaction, to offer proof-of-concept experiments with modest sample sizes not unlike those described in this thesis (e.g., Brinkman, Broekens, & Heylen, 2015).

Several reasons account for the sample sizes used in this thesis. Firstly, small samples were a result of resource limitations and other practical constraints that go hand-in-hand with doing research at a graduate student level (e.g., laboratory access, participant compensation, and so on). Second, obtaining large samples for studies of the kind presented in this thesis is made difficult by the fact that cyranic interactions require the time and effort of speech shadowing confederates who must first take time to master speech shadowing and then learn to perform in accordance with their role in a consistent manner that maintains the integrity of the experimental interactions and establishes uniformity across trials. Thirdly, experiments such as those performed for this thesis require one-on-one interactions followed by extensive debrief interviews conducted by the researcher, which negates the possibility of running parallel trials or conducting experiments remotely (e.g., by Amazon Mechanical Turk). If the cyranoid method goes on to be adopted by researchers in other laboratories, it is hoped that efficiencies are found that might enable more streamlined interactions and larger samples without sacrifices to the richness of the interactions as experienced by participants.

2. Technological constraints. Technological constraints placed various limitations on the

cyranic interactions explored in this thesis. In the case of cyranoids featuring human sources (e.g., those explored in Chapter 2 and Chapter 5), sources were limited to receiving audio and video stimuli of the interactant through a computer monitor from their remote location and therefore did not literally experience “fully being” the mind of the shadower (in terms of receiving the full range of stimuli that accompany in person social interaction). This thesis did not explore whether and to what extent this affected how sources communicated. Interestingly, however, there are computer scientists who are developing technologies that allow a remote person to more fully experience the fuller range of sensory information present to a remote person (discussed below in the section “Going forward: Future research areas”), and these technologies might allow sources more complete immersion into the worlds of their shadowers. Similarly, the agent-sources within the echoborgs explored in this thesis were constricted to receiving textual input by-way-of the minimal technological dependency format of human-echoborg interaction. It is highly plausible that had these agents had access to more forms of social input (e.g., the ability to read facial cues and other body language), the social performance of the echoborgs (as perceived by interactants) would have been less abnormal and more contextually fine-tuned. There presently exists an array of sensory technologies for artificial agents that can read and interpret visual and audio data (e.g., the facial expressions of a human interlocutor; see Gruebler, Berenz, & Suzuki, 2011), and these technologies have been demonstrated with devices such as the fully autonomous androids discussed in Chapter 3. It is not unrealistic to imagine these sensory technologies being incorporated into future iterations of the echoborg method, but testing such contraptions fell outside the purview of this thesis.

What this particular category of limitations points toward is the notion of “presence” (Cummings & Bailenson, 2015), a construct that encompasses the sense of “being there” offered by a particular form of technologically mediated communication. The sense of presence experienced by interactants who encounter a cyranoid or echoborg is arguably quite high given that interactants encounter a fleshy human body in person. The sources (human and agent) in this thesis, on the other hand, did not experience this same degree of presence given the devices through which they remotely communicated.

3. Replication issues. Experimental social psychology is presently in somewhat of a crisis as many high-profile studies, particularly those that involve priming as a means of implementing a social stimulus, have failed to be replicated (see Open Science Collaboration,

2015). This has reinvigorated the longstanding debate about certain structural problems facing the discipline, such as the drastic publication bias in favor of non-null findings, that while not unique to the discipline, are certainly more pronounced in social psychology relative to other fields within academia (Yong 2012a, 2012b). It is worth, then, discussing the replicability of the studies detailed in this thesis as well as replication challenges inherent in the cyranoid method.

Though the designs and general parameters of the experiments conducted for this thesis can certainly be replicated by other researchers, each cyranoid is a unique stimulus that cannot always be redeployed in future research. Cyranoids have as their component parts actual human beings (sources and shadowers), and though general features of each source and shadower can be found in other people who volunteer as cyranoids in other research laboratories (e.g., gender, age, race, appearance, and so on), certain physical and dispositional idiosyncrasies of any one source or shadower can never be exactly imitated by another person. The overall personas of the cyranoids and echoborgs featured in this thesis, therefore, can probably never again be exactly replicated elsewhere. On the other hand, it is arguable that conversational agent sources encompass a highly replicable stimulus as they are guaranteed to perform in a consistent and unbiased manner in all potential research iterations insofar as no changes are made to their source code or to the corpus of data from which they formulate responses.

Each of the experiments described in this thesis featured non-scripted encounters (i.e., participants were not given any instruction to speak in a certain way or within certain parameters). Many of the experiments, in fact, were designed to invoke the mundane characteristics of in person encounters between strangers wherein neither party is given any information about what to expect from the other. Interactants were encouraged to speak about whatever they wanted. This style of experimental interaction is slightly atypical from what one tends to see in social psychological research, wherein to assist replicability (among other things) participants are usually given a task to perform (e.g., a competitive game or collaborative activity) that is constrained by scripts, rules, and roles delved out to each interlocutor. The point of exploring unscripted encounters was to see how conversations between interactants and cyranoids/echoborgs would naturally evolve and, importantly, to reduce interactants' sense that the interactions might be scripted. The designs of these experiments therefore constituted a tradeoff between high realism and high replicability,

favoring high realism.

4. Control problems. Cyranoid experiments of the kind described in this thesis are inherently challenging to control and thereby completely de-confound. As argued at length in Chapter 2, one of the more important elements that *can* be controlled is the independent manipulation of the source and shadower, which allows one to study how the same body is perceived when given various minds (i.e., sources), and how the same source is perceived when paired with different bodies (i.e., shadowers). What are more difficult to control, however, are various idiosyncrasies that arise from trial to trial, especially in unscripted interaction scenarios. It is important to not think of the shadower as an entity being entirely controlled by a source; a shadower is not a puppet. Shadowers experience reflexive, embodied reactions to cyranic interaction situations (e.g., blushing, urges to laugh, and so on) that can be anticipated but never fully suppressed. The full range of social cues exhibited by the shadower, therefore, can only be made consistent from trial to trial to a certain extent. To account for this, the experiments in this thesis involved extensive piloting and rehearsal such that shadowers could familiarize themselves with the spontaneous sensations that arise while covertly shadowing and develop a means of making their body language *fit* the social situation to as normal-a-degree as possible.

The various amalgamations of gadgetry used to create the cyranoids and echoborgs described in this thesis brought with them certain control challenges that in some cases amplified those inherent to speech shadowing. For example, the audio-relay delays seen with the echoborg studies compounded the shadower body language consistency issues. When one is covertly shadowing for a conversational agent that is by its very nature deviant in terms of conversational capacity, it is quite difficult to maintain a consistent bodily demeanor during awkward and unnatural silences. This issue no doubt affected interactants' perceptions, as many of them commented on their interlocutor's unnatural delays in speech, so it remains to be seen how perception of a conversational agent given a human body to speak through is perceived when turn-taking latencies are normalized. Another confound brought about by the gadgetry used in this thesis involves the radio relay apparatus via which sources spoke to their shadowers. At times the radio signal became weak, leading to shadowing hesitations and in some cases mild inaccuracies. Best attempts were made to optimize audio clarity for the shadower, but at random times audio clarity became poor and whatever affect this had on interactants' perceptions of the cyranoid/echoborg went unaccounted for.

Going forward: Ideas for future research

Improving immersion and presence. One thought for how to augment the sense of presence experienced by the source within a cyranoid would be to make use of immersive audio/visual technology such as the Oculus Rift or similar IVET hardware. Primarily designed for virtual reality, these systems grant users a 3D experience while blocking out audio/visual stimuli from the wearer's surrounding environment. One could imagine equipping a source with such a device in order to accomplish fuller immersion into the experience of the shadower. In their meta-analytic review of the effectiveness of various features of immersive technology on enhancing presence, Cummings and Bailenson (2015) discuss the importance of three features in particular: (1) stereoscopic visuals (i.e., the user of an immersive device should encounter visuals that recreate the depth and three-dimensionality of real world experience), (2) wide fields-of-view (i.e., the user of an immersive device should be able to experience a wide field-of-vision much like what they experience in the real world), and (3) high degrees of user-tracking (i.e., the amount of behaviors that are tracked by the immersive technology, which is not so important when it comes to the cyranoid method given the fact that sources will be looking at high-quality videos of an interactant rather than the virtual representation of an interactant). The way forward for cyranoid research, therefore, may be to involve immersive technology capable of stereoscopic visuals and a wide fields-of-view.

Another way forward for increasing the presence felt by sources involves the use of haptic devices that allow the source to experience tactile sensations felt by the shadower. There are many proof-of-concept and early design demonstrations of haptic devices being incorporated into robotic social tele-presence systems, such as gloves that allow a source (i.e., a tele-operator) to sense the hand sensations felt by a remote robotic device (e.g., Nakanishi, Tanaka, & Wada, 2014). It is conceivable that such technologies could further bring the source into the sensory world of the shadower. In future iterations of the cyranoid method, it would be well worth incorporating haptic devices into cyranic contraptions, though it is most likely that in such situations it would be difficult to construct a covert cyranoid encounter as the shadower would have to wear sensory devices on their body that linked with the haptic devices felt by the source.

Varying the degree of shadower agency and the “partial echoborg.” In the experiments conducted for this thesis, as well as those carried out by Milgram, shadowers were instructed to not interject with any of their own thoughts while communicating with interactants. What

has yet to be studied, however, are scenarios in which shadowers speak to interactants more or less freely with their own thoughts yet can defer to a source (or sources) in particular instances. Milgram alluded to this possibility in his grant applications yet never produced such a study. This concept is intriguing because it introduces a shadower's true conversational disposition to a cyranic interaction and supplements it with the disposition of one or potentially many other dispositions.

A particularly intriguing permutation of this idea is the “partial echoborg”: a person paired with an agent to whom they could defer as a source when necessary. Imagine going about life paired with advanced machine intelligence that verbally feeds you information relevant to your social context that you could then utter as if the information had originated within your own mind. A sufficiently advanced form of machine intelligence would be able to scan the Internet and information databases in real-time, granting you the ability to answer practically any answerable question posed to you. Armed with little more than a discreet earpiece, you, as a partial echoborg, could achieve superhuman levels of information recall (imagine, for instance, the agent to whom you were paired was a personalized version of IBM’s Watson program). The echoborg studies described in the current thesis could serve as the basis for future proof-of-concept research into this idea using artificial agent technologies that are currently available.

Far from being merely the stuff of science fiction, the partial echoborg concept can potentially see early fruition as a means of enhancing the livelihood and sociability of people with certain cognitive impairments, such as dementia. People who suffer from cognitive decline generally have difficulties with information recall and remembering the intermediary steps involved in completing rudimentary tasks. There currently exists a class of artificial agents and other technologies known as “cognitive prosthetics” that are designed to interface with individuals suffering from cognitive decline such that they can better overcome impaired recall and achieve better routine functionality (for a review of such technologies, see Gillespie, Best, & O’Neil, 2012). O’Neill and Gillespie (2008) argue that many of these cognitive prosthetics rely on visual interfaces that are not intuitive to a person suffering from cognitive decline and that such technologies often further complicate or disorient the person in need of assistance. The authors suggest that auditory interfaces are preferable to visual interfaces and give examples of auditory assistive technologies that have been shown to help scaffold users’ functioning. The current iterations of these auditory interfaces act as

interlocutors – systems that, by asking questions and scaffolding conversation, guide a user toward a certain objective (e.g., preparing a meal). The echoborg variant of this scenario would be to develop inner-ear auditory cognitive prostheses that act as *first-person* systems (i.e., rather than function as a user’s interlocutor, such systems would actually prompt a user with what they could or should say in a particular situation, such as recalling the name of their doctor or an address).

The cyranoid method as a means of augmenting cognitive capacity contributes to the domain of social science and technology research known as “transhumanism,” in part a cultural and intellectual movement that advocates radically altering human cognitive and physical capacities through merging with technology so as to enhance humanity’s ability to operate in the current world as well as ensure the long-term survival of the species (see Bostrom, 2005). In this sense, echoborg research can serve as a window into a potentially not-too-distant future wherein wearable devices that assist with, among other things, first-person information recall are ubiquitous and *necessary* for mundane social interaction. And because the cyranoid method preserves the human interface (i.e., the technology that enables a cyranic interaction is hidden from the view of interactants), the echoborg concept as a route to transhuman cognitive capacity might do better in being adopted by people in need or want of enhanced cognitive capacity than other types of technological augmentation that are more overt.

Multiple sources? Multiple shadowers? Chapter 2 and Chapter 5 alluded to possible variants of the basic tripartite structure of cyranic interactions. These include situations wherein multiple sources shadow for a single shadower (a context explored by Milgram and his research assistants in their early piloting of the method) as well as situations wherein a single source provides words and directives to multiple shadowers. The former of these scenarios is interesting because one could imagine performing a research study that involved assessing to what extent interactants perceive a stable disposition in an interlocutor who is in reality covertly shadowing simultaneously or successively for varied sources. The later scenario seems a candidate for a breeching experiment wherein an interactant would confront multiple individuals performing as covert shadowers and have the experience of talking to one person (a single source) simultaneously through different people. Though such scenarios are arguably more theatrical in nature than useful social psychologically, they can potentially be incorporated into activities that do produce some sort of positive psychological effect (e.g., as classroom activities for schoolchildren).

“Cyrafour” and other concept studies in interaction design. As part of their ongoing work in piloting creative forms of technology-mediated interaction, Encinas and Mitchell (2015) have explored a novel cyranic scenario that they call “cyrafour.” The concept involves speech shadowers duplicating the speech of two remote interlocutors in dialog with one another. With cyrafour, witnesses observe two individuals speaking to each other face-to-face when in reality these two individuals are merely shadowing an identical conversation being had by unseen sources. The authors suggest that such an activity can serve as a role-switching exercise. Though perhaps more of a dramatic activity than a scholarly one, cyrafour is an example of the cyranoid method inspiring permutations in experiential, first-person forms of research.

Indeed, the multidisciplinary field of interaction design is where the cyranoid method can see wide future application. Interaction design brings together psychologists, computer scientists, engineers, software developers, and other scholars who aim to develop new means of socializing with and through technology. Interaction design is more interested in exploring ways of human-human and human-computer interaction that are *possible* even if not always practical or economically viable (Cooper, Reimann, & Cronin, 2007). The agenda of many researchers within the field of interaction design is to provide proof-of-concept manifestations of novel interaction methods such that they may inspire real-world applications or, at the very least, reveal something interesting about human behavior and/or human social interaction. The cyranoid method, therefore, is a natural fit within the world of interaction design, and the variations of the method that have been covered in this thesis are merely a few of the potentially vast number of design permutations the method can take within the discipline. The echoborg in particular holds vast potential within interaction design given that at its core the discipline seeks to build new forms of human-technology interaction.

Final Remarks

Milgram’s investigatory style was fundamentally a first-principles approach to studying human sociality. He started every research endeavor by simply asking what it is we really *know* about a type of human behavior and how can we build methods from the ground up that help us to more fully understand the function of that behavior in a particular social context. Milgram sought to explore, even if only dramaturgically, the bedrock social psychological phenomena that govern how we operate with one another (Blass, 2004). His cyranoid method was derivative of such an approach, as have been the variants of it explored in this thesis. So

what *do* we know? We know that we reflexively attribute verbal autonomy to people we encounter. We know that we tend to ascribe authorship of the words we notice other people speaking to the brains inside those people's heads. And we know that characteristics of the body – the interface – largely determine how we perceive and orient to these words. The cyranoid method provides us with both a technique with which to more fully understand these tendencies as well as a metaphor that encapsulates something else social psychologists have come to intuitively know, that all of us, much of the time, articulate the ideas of other people (see Bakhtin, 2010; Maybin, 2001). People we talk to, people whose work we read or witness on television or online, people who are in positions of authority relative to us... all people who, for the most part, are themselves speaking ideas that originated elsewhere. This isn't to say that humans are incapable of articulating original thoughts, as we most certainly do, but only that behind so much of what we say and what we notice other people say is a hidden genealogy that we are largely unaware of yet which connects us in a fundamental way to our fellow human beings both past and present. We become who we are, both in terms of how we see ourselves and how we are seen by others, through a cyranic process wherein we continually channel the ideas and opinions of myriad sources who themselves are or were the result of a cyranic process. Whenever we rearticulate an idea we heard from someone or somewhere else, our bodies are in that moment merely the most recent interface through which that idea makes contact with the social world. This is a truly bizarre situation.

It is fair to say that even conversational agents, such as those used in this thesis's echoborg experiments, are in some sense cyranoids, as the content of what they articulate and the rules that govern what they articulate are determined by software engineers and the humans with whom they interact. The social behaviors of androids, robots, chat bots, and other forms of artificial intelligence are governed by code whose origins lay external to these systems in the minds of human beings. They are hostage to this code in the same way Milgram imagined people are when in the midst of overwhelming authority – they exist in a perpetual agentic state. So the fact that we see these technologies as somehow different from us is revelatory of a "hardware bias" that we seem to suffer from. We tend to make distinctions about people and things based on how they appear to us to be rather than see the causal chain pulling the strings that exists behind the interface. So insofar as chat bots, robots, androids, and other social machines are merely interfaces for human-derived communication in the same way that we are, these technologies *are* in some sense human. The covert echoborg studies show that it is probable that once machine intelligence hardware is sufficiently advanced so as to

allow for interactions with us on a level playing field (i.e., when their interfaces elicit the same social psychological expectations of human-human interaction), the category we know as “human” will forever be altered.

For all his shortcomings, Milgram had great intuitions as to how to illuminate these fascinating concepts. His entire body of work, from the obedience studies through to his cyranoid pilots, attests to a curiosity in those social processes that are forever present yet not always immediately obvious. Novel methods, now matter how peculiar, were how he pulled back the curtain on these processes, and the cyranoid method is perhaps Milgram’s *most* peculiar invention. So peculiar that those outside his small circle dismissed the idea during his lifetime, and so peculiar that it for the better part of three decades was nothing more than an easily forgotten footnote in the history of social psychology. But in Milgram’s own words, “certain research methods are useful for the very reason that they alter the ordinary manner in which things work; by doing so they force us into a more rigorous examination of the processes of normal functioning” (Milgram, 1979, p. 1). Here at the culmination of this thesis, which started as a simple attempt to replicate Milgram’s cyranoid idea and morphed into a new frontier of human-agent interaction and first-person research, this statement of Milgram’s, above all else, has proven true. The cyranoid method gives us a fascinating window into how the body, or interface, becomes a prism through which words become refracted. It provides us with a means of stepping into the shoes of other individuals so as to experience to some degree the world that they live in and in doing so become more self-aware. It gives us a way of interacting through a truly human interface with technology. The method really does give us a means of digging more deeply into those social psychological phenomena that are not just interesting from a scientific or academic perspective, but that reverberate with the experiences and imaginations of ordinary people.

REFERENCES

- Abildgaard, J.R., & Scharfe, H. (2012). A geminoid as a lecturer. In S.S. Ge, O. Khatib, J. Cabibihan, R. Simmons, & M. Williams (Eds.), *Social robotics: Proceedings of the 4th International Conference on Social Robotics* (pp. 408-417). Berlin, Germany: Springer.
- Adair, J.G. (1991). Social cognition, artifact, and the passing of the so-called crisis in social psychology. *Canadian Psychology*, 32(3), 445-450.
- Albert, S. (1983, November). [Letter addressed to S. Milgram]. Stanley Milgram Papers Archive Manuscripts and Archives Department, Sterling Memorial Library, Yale University, New Haven, CT.
- Allport, G. (1985). The historical background of social psychology. In G. Lindzey & E. Aronson (Eds.), *Handbook of social Psychology* (3rd ed., Vol. 1, pp.1-46). New York: Random House.
- Almereyda, M. (Writer/Director). (2015). *Experimenter* [Motion picture]. United States: BB Film Productions.
- Alterman, R. (2007). Representation, interaction, and intersubjectivity. *Cognitive Science*, 31(5), 815-841.
- American Psychological Association (1973). Ethical principles in the conduct of research with human participants. *American Psychologist*, 28(1), 79-80.
- American Psychological Association (2010). *Ethical principles of psychologists and code of conduct*. Washington, DC: American Psychological Association. Retrieved from: <http://www.apa.org/ethics/code/principles.pdf>
- Anderson, L. (2006). Analytic autoethnography. *Journal of Contemporary Ethnography*, 35(4), 373-395.
- Arundale, R.B. (2010). Constituting face in conversation: Face, facework, and interaction achievement. *Journal of Pragmatics*, 42(8), 2078-2105.
- Asch, S.E. (1951). *Effects of group pressure upon the modification and distortion of*

judgment. In H. Guetzkow (ed.), *Groups, leadership and men*. Pittsburgh: Carnegie Press.

Bailenson, J.N., Beall, A.C., Loomis, J., Blascovich, J., & Turk, M. (2004). Transformed social interaction: Decoupling representation from behavior and form in collaborative virtual environments. *Presence, 13*(4), 428-441.

Bailenson, J.N., Blascovich, J., & Guadagno, R.E. (2008). Self-representations in immersive virtual environments. *Journal of Applied Social Psychology, 38*(11), 2673-2690.

Bailenson, J.N., Swinth, K., Hoyt, C., Persky, S., Dimov, A., & Blascovich, J. (2005). The independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments. *Presence, 14*(4), 379-393.

Bailenson, J.N., & Yee, N. (2005). Digital chameleons: Automatic assimilation of nonverbal gestures in immersive virtual environments. *Psychological Science, 16*(10), 814-815.

Bailenson, J.N., Yee, N., Blascovich, J.J., Guadagno, R.E. (2008). Transformed social interaction in mediated interpersonal communication. In E. Konijn, M. Tanis, S. Utz, & A. Linden (Eds.), *Mediated interpersonal communication* (pp. 77-99). Mahwah: Lawrence Erlbaum.

Bailenson, J.N., Yee, N., Patel, K., & Beall, A.C. (2008). Detecting digital chameleons. *Computers in Human Behavior, 24*(1), 66-87.

Bailly, G. (2003). Close shadowing natural versus synthetic speech. *International Journal of Speech Technology, 6*(1), 11-19.

Balzarotti, S., Piccini, L., Andreoni, G., & Ciceri, R. (2014). “I know that you know how I feel”: Behavioral and physiological signals demonstrate emotional attunement while interacting with a computer simulating emotional intelligence. *Journal of Nonverbal Behavior, 38*(3), 283-299.

Bakhtin, M.M. (2010). *Speech genres and other late essays*. Austin: University of Texas Press.

Bänziger, T., Grandjean, D., & Scherer, K.R. (2009). Emotion recognition from expressions in face, voice, and body: The multimodal emotion recognition test (MERT). *Emotion*

9(5), 691–704.

Barker, V., Giles, H., & Harwood, J. (2004). Inter- and intragroup perspectives on intergenerational communication. In J.F. Nussbaum & J. Coupland (Eds.), *Handbook of communication and aging research* (2nd ed.). Mahwah: Lawrence Erlbaum Associates.

Baumeister, R., Vohs, K.D., & Funder, D.C. (2007). Psychology as a science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*, 2(4), 396-403.

Baumrind, D. (1964). Some thoughts on ethics of research: After reading Milgram's "Behavioral Study of Obedience." *American Psychologist*, 19(6), 421-423.

Becker, H.S., & Geer, B. (1957). Participant observation and interviewing: A comparison. *Human Organization*, 16(3), 28-32.

Becker-Asano, C., & Ishiguro, H. (2011). Intercultural differences in decoding facial expressions of the android robot Geminoid F. *Journal of Artificial Intelligence and Soft Computing Research*, 1(3), 215-231.

Becker-Asano, C., Ogawa, K., Nishio, S., & Ishiguro, H. (2010). Exploring the uncanny valley with Geminoid HI-1 in a real-world application. In K. Blashki (Ed.), *Proceedings of the IADIS International Conference on Interfaces and Human Computer Interaction* (pp. 121-128). Lisbon, Portugal: IADIS Press.

Berk, L. E., & Winsler, A. (1995). *Scaffolding children's learning: Vygotsky and early childhood education*. Washington, DC: National Association for the Education of Young Children.

Berry, D.S., & McArthur, L.Z. (1986). Perceiving character in faces: The impact of age-related craniofacial changes on social perception. *Psychological Bulletin*, 100(1), 3-18.

Bessho, F., Harada, T., & Kuniyoshi, Y. (2012). Dialog system using real-time crowdsourcing and Twitter large-scale corpus. In G.G. Lee, J. Ginzburg, C. Gardent, & A. Stent (Eds.), *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialog* (pp. 227-231). Stroudsburg: Association for

Computational Linguistics.

Bessiere, K., Seay, A.F., & Kiesler, S. (2007). The ideal elf: Identity exploration in World of Warcraft. *CyberPsychology & Behavior, 10*(4), 530-535.

Biklen, S., & Moseley, C. (1988). "Are you retarded?" "No, I'm Catholic": Qualitative methods in the study of people with severe handicaps. *Journal of the Association of Severe Handicaps, 13*, 155-162.

Blascovich, J., Loomis, J., Beall, A.C., Swinth, K.R., Hoyt, C.L., & Bailenson, J.N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry, 12*(2), 103-124.

Blass, T. (2004). *The man who shocked the world: The life and legacy of Stanley Milgram*. New York: Basic Books.

Boellstorff, T. (2008). *Coming of age in Second Life: An anthropologist explores the virtually human*. Princeton: Princeton University Press.

Bohner, G., Chaiken, S., & Hunyadi, P. (1994). The role of mood and message ambiguity in the interplay of heuristic and systematic processing. *European Journal of Social Psychology, 24*(1), 207-221.

Booth, T., & Booth, W. (1996). Sounds of silence: Narrative research with inarticulate subjects. *Disability & Society, 11*(1), 55-70.

Boring, E.G. (1915). The sensations of the alimentary canal. *American Journal of Psychology, 26*(1), 1-57.

Bostrom, N. (2005). A history of transhumanist thought. *Journal of Evolution and Technology, 14*(1), 1-25.

Bouchard, T.J., Jr. (1976). Field research methods: Interviewing, questionnaires, participant observation, systematic observation, unobtrusive measures. In M.D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 363-413). Chicago: Rand McNally.

Branigan, H.P., Pickering, M.J., Pearson, J., McLean, J.F., & Brown, A. (2011). The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers.

Cognition, 121(1), 41-57.

Brewer, M.B., Dull, V., & Lui, L.N. (1981). Perceptions of the elderly: Stereotypes and prototypes. *Journal of Personality and Social Psychology*, 41(4), 656-670.

Brewer, M.B., & Lui, L.N. (1989). The primacy of age and sex in the structure of person categories. *Social Cognition*, 7(3), 262-274.

Brinkman, W., Broekens, J., & Heylen, D. (Eds.). (2015). *Proceedings of the 15th International Conference on Intelligent Virtual Agents (IVA 2015)*. Cham, Switzerland: Springer.

British Psychological Society (2000). *Code of conduct, ethical principles, and guidelines*. Leicester, UK: British Psychological Society.

Brave, S., Nass, C., & Hutchinson, K. (2005). Computers that care: Investigating the effects of orientation of emotion exhibited by an embodied computer agent. *International Journal of Human-Computer Studies*, 62(2), 161-178.

Brennan, S.E. (1991). Conversation with and through computers. *User Modeling and User-Adapted Interaction*, 1(1), 67-86.

British Psychological Society (2009). *Code of ethics and conduct*. Leicester, UK: British Psychological Society. Retrieved from:

http://www.bps.org.uk/system/files/documents/code_of_ethics_and_conduct.pdf

Cabanac, M. (2004). Dionysians and Apollonians [Peer commentary on “Self-experimentation as a source of new ideas: Ten examples about sleep, mood, health, and weight,” by S. Roberts]. *Behavioral and Brain Sciences*, 27(2), 263-264.

Cantor, N., & Mischel, W. (1979). Prototypes in person perception. *Advances in Experimental Social Psychology*, 12, 3-52.

Carpenter, R. (2014). *Cleverbot* [computer program]. Retrieved from
<http://www.cleverbot.com>.

Carpenter, R. (2015). *Cleverbot* [computer program]. Retrieved from
<http://www.cleverbot.com>.

Cassell, J. (2000). Embodied conversational interface agents. *Communications of the ACM*,

43(4), 70-78.

Cassell, J., Sullivan, J., Prevost, S., & Churchill, E. (Eds.). (2000). *Embodied conversational agents*. Cambridge: MIT Press.

Cassell, J., & Tartaro, A. (2007). Intersubjectivity in human-agent interaction. *Interaction Studies*, 8(3), 391-410.

Cassie, R., & Gibson, C.L. (1997). The effectiveness of repair strategies used by people with hearing losses and their conversational partners. *Volta Review*, 99(4), 203-218.

Cesario J., & Higgins, E.T. (2008). Making message recipients “feel right”: How nonverbal cues can increase persuasion. *Psychological Science*, 19(5), 415-420.

Chaiken, S. (1979). Communicator physical attractiveness and persuasion. *Journal of Personality and Social Psychology*, 37(8), 1387-1397.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, 39(5), 752-766.

Chaiken, S., & Maheswaran, D. (1994). Heuristic processing can bias systematic processing: Effects of source credibility, argument ambiguity, and task importance on attitude judgment. *Journal of Personality and Social Psychology*, 66(3), 460-473.

Chaminade, T., Rosset, D., Da Fonseca, D., Nazarian, B., Lutcher, E., Cheng, G., & Deruelle, C. (2012). How do we think machines think? An fMRI study of alleged competition with an artificial intelligence. *Frontiers in Human Neuroscience*, 6(103), 1-9.

Chang, H. (2008). *Autoethnography as method*. Walnut Creek: Left Coast Press.

Chen, M., & Bargh, J.A. (1997). Nonconscious behavioral confirmation processes: The self-fulfilling consequences of automatic stereotype activation. *Journal of Experimental Social Psychology*, 33(5), 541-560.

Chen, S., & Chaiken, S. (1999). The heuristic-systematic model in its broader context. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 73-96). New York: The Guilford Press.

Chomsky, N. (2008). Turing on the “imitation game.” In R. Epstein, G. Roberts, & G. Beber

- (Eds.), *Parsing the Turing Test: Philosophical and methodological issues in the quest for the thinking computer* (pp. 103-106). New York: Springer.
- Clark, H.H. (2009). *Using language*. Cambridge, UK: Cambridge University Press.
- Clark, H.H., & Schaefer, E.F. (1987). Collaborating on contributions to conversations. *Language and Cognitive Processes*, 2(1), 19-41.
- Clark, H.H., & Schaefer, E.F. (1989). Contributing to discourse. *Cognitive Science*, 13(2), 259-294.
- Clark, H.H., & Brennan, S.E. (1991). Grounding in communication. In L.B. Resnick, J.M. Levine, & S.D. Teasley (Eds.), *Perspectives on Socially Shared Cognition* (pp. 222-233). Washington, DC: American Psychological Association.
- Condon, W.S., & Ogston, W.D. (1966). Sound film analysis of normal and pathological behavior patterns. *Journal of Nervous and Mental Disease*, 143(3), 338-347.
- Cooper, A., Reimann, R., & Cronin, D. (2007). *About face 3: The essentials of interaction design*. Indianapolis: Wiley.
- Copeland, B.J. (2000). The Turing Test*. *Minds and Machines*, 10(4), 519-539.
- Cornish, F., Zittoun, T., & Gillespie, A. (2007). A cultural psychological reflection on collaborative research. *Forum: Qualitative Social Research*, 8(3). Retrieved from <http://www.qualitative-research.net/index.php/fqs/article/view/309/677>.
- Corti, K., & Gillespie, A. (2015). A truly human interface: Interacting face-to-face with someone whose words are determined by a computer program. *Frontiers in Psychology*, 6(634), 1-18.
- Corti, K., & Gillespie, A. (2015). Revisiting Milgram's cyranoid method: Experimenting with hybrid human agents. *Journal of Social Psychology*, 155(1), 30-56.
- Corti, K., & Gillespie, A. (2015, May 11). *The "echoborg method" of human-agent interaction* [Video file]. Retrieved from <https://www.youtube.com/watch?v=NtWLCZZYM64>
- Corti, K., Reddy, G., Choi, E., & Gillespie, A. (2015). The researcher as experimental subject: Using self-experimentation to access experiences, understand social

phenomena, and stimulate reflexivity. *Integrative Psychological and Behavioral Science*, 49(2), 288-308.

Cummings, J.J., & Bailenson, J.N. (2015). How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychology* (advance online publication), 1-38.

Dahlbäck, N., Jönsson, A., & Ahrenberg, L. (1993). Wizard of Oz studies: Why and how. In W.D. Gray, W.E. Hefley, & D. Murray (Eds.), *Proceedings of the 1st International Conference on Intelligent User Interfaces* (pp. 193-200). New York: ACM.

Danziger, K. (1979). The positivist repudiation of Wundt. *Journal of the History of the Behavioral Sciences*, 15(3), 205-230.

Danziger, K. (1980). Wundt's psychological experiment in the light of his philosophy of science. *Psychological Researcher*, 42, 109-122.

Danziger, K. (1985). The origins of the psychological experiment as a social institution. *American Psychologist*, 40(2), 133-140.

Danziger, K. (1990). *Constructing the subject: Historical origins of psychological research*. Cambridge, UK: Cambridge University Press.

Danziger, K. (2000). Making social psychology experimental: A conceptual history, 1920-1970. *Journal of the History of the Behavioral Sciences*, 36(4), 329-347.

Dambrun, M., & Vatiné, E. (2010). Reopening the study of extreme social behaviors: Obedience to authority within an immersive video environment. *European Journal of Social Psychology*, 40(5), 760-773.

Dautenhahn, K. (2007). Socially intelligent robots: Dimensions of human-robot interaction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480), 679-704.

De Melo, C.M., Carnevale, P., & Gratch, J. (2011). The impact of emotion displays in embodied agents on emergence of cooperation with people. *Presence*, 20(5), 449-465.

De Preester, H. (2008). From *ego* to *alter ego*: Husserl, Merleau-Ponty and a layered approach to intersubjectivity. *Phenomenology and the Cognitive Sciences*, 7(1), 133-

- Dennett, D.C. (2004). Can machines think? In C. Teuscher (Ed.), *Alan Turing: Life and legacy of a great thinker* (pp. 295-316). Berlin, Germany: Springer.
- Dennett, D.C. (2007). Heterophenomenology reconsidered. *Phenomenology and the Cognitive Sciences*, 6(1), 247-270.
- Denzin, N.K., & Lincoln, Y. (2000). *Handbook of qualitative research*. Thousand Oaks: Sage Publications.
- Devine, P.G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56(1), 5-18.
- Dick, P.K. (1968). *Do androids dream of electric sheep?* New York: Doubleday.
- DiFonzo, N., Hantula, D.A., & Bordia, P. (1998). Microworlds for experimental research: Having your (control and collection) cake, and realism too. *Behavior Research Methods*, 30(2), 278-286.
- Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, 24(3), 285-290.
- Doherty, A.R., Caprani, N., Conaire, C.Ó., Kalnikaite, V., Gurrin, C., Smeaton, A.F., & O'Connor, N.E. (2011). Passively recognising human activities through lifelogging. *Computers in Human Behavior*, 27(5), 1948-1958.
- Donath, J.S. (1998). Identity and deception in the virtual community. In P. Kollock & M. Smith (Eds.), *Communities in cyberspace*. London, UK: Routledge.
- Dotsch, R., & Wigboldus, D.H.J. (2008). Virtual prejudice. *Journal of Experimental Social Psychology*, 44(4), 1194-1198.
- Dougherty, E.G., & Scharfe, H. (2011). In B. Mutlu, C. Bartneck, J. Ham, V. Evers, & T. Kanda (Eds.), *Social robotics: Proceedings of the 3rd International Conference on Social Robotics* (pp. 95-103). Berlin, Germany: Springer.
- Drew, P. (1997). “Open” class repair initiators in response to sequential sources of troubles in conversation. *Journal of Pragmatics*, 28(1), 69-101.

- Dumitrica, D., & Gaden, G. (2009). Knee-high boots and six-pack abs: Auto ethnographic reflections on gender and technology in Second Life. *Journal of Virtual Worlds Research*, 1(3), 3-23.
- Duncan, B.L. (1976). Differential social perception and attribution of intergroup violence: Testing the lower limits of stereotyping of blacks. *Journal of Personality and Social Psychology*, 34(4), 590-598.
- Eagley, A.H., Ashmore, R.D., Makhijani, M.G., & Longo, L.C. (1991). What is beautiful is good, but...: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, 110(1), 109-128.
- Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6(3-4), 169-200.
- Elms, A.C. (1975). The crisis of confidence in social psychology. *American Psychologist*, 30(10), 967-976.
- Encinas, E., & Mitchell, R. (2015). Cyrafour.
- Farr, R.M. (1978). On the social significance of artifacts in experimenting. *British Journal of Social and Clinical Psychology*, 17(4), 299-306.
- Farr, R.M. (1983). Wilhelm Wundt (1832-1920) and the origins of psychology as an experimental and social science. *British Journal of Social Psychology*, 22(4), 289-301.
- Farr, R.M. (1996). *The roots of modern social psychology*. Oxford, UK: Blackwell.
- Feil-Seifer, D., Skinner, K., & Matarić, M.J. (2007). Benchmarks for evaluating social assistive robotics. *Interaction Studies*, 8(3), 423-439.
- Field, A. (2009). *Discovering statistics using SPSS*. London, UK: Sage Publications.
- Fiske, S.T. (2010). Interpersonal stratification: Status, power, and subordination. In S.T. Fiske, D.T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology*. Hoboken: John Wiley & Sons.
- Fleck, R., & Fitzpatrick, G. (2009). Teachers' and tutors' social reflection around SenseCam images. *International Journal of Human-Computer Studies*, 67(12), 1024-1036.

- Fowler, C.A., Brown, J.M., Sabadini, L., & Weihing, J. (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language*, 49(3), 396-413.
- Fox, J., & Bailenson, J.N. (2009). Virtual self-modeling: The effects of vicarious reinforcement and identification on exercise behaviors. *Media Psychology*, 12(1), 1-25.
- Fox, J., Bailenson, J.N., & Tricase, L. (2013). The embodiment of sexualized virtual selves: The Proteus effect and experiences of self-objectification via avatars. *Computers in Human Behavior*, 29(3), 930-938.
- Frank, M.G., & Ekman, P. (1997). The ability to detect deceit generalizes across different types of high-stake lies. *Journal of Personality and Social Psychology*, 72(6), 1429-1439.
- French, R.M. (2000). The Turin Test: The first 50 years. *Trends in Cognitive Sciences*, 4(3), 115-122.
- French, R.M. (2012). Moving beyond the Turing Test. *Communications of the ACM*, 55(12), 74-77.
- Friesen, N. (2009). Discursive psychology and educational technology: Beyond the cognitive revolution. *Mind, Culture, and Activity*, 16(2), 130-144.
- Frohlich, D., Drew, P., & Monk, A. (1994). Management of repair in human-computer interaction. *Human-Computer Interaction*, 9(3-4), 385-425.
- Gabriel, P. (1986). We do what we're told (Milgram's 37) [Recorded by P. Gabriel]. On *So* [Music album]. Bath, UK: Charisma/Geffen.
- Gallagher, H.L., Jack, A.I., Roepstorff, A., & Frith, C.D. (2002). Imaging the intentional stance in a competitive game. *NeuroImage*, 16(3), 814-824.
- Gallagher, S., & Sørensen, J.B. (2006). Experimenting with phenomenology. *Consciousness and Cognition*, 15(1), 119-134.
- Garfinkel, H. (1964). Studies of the routine grounds of everyday activities. *Social Problems*, 11(3), 225-250.

- Garfinkel, H. (1967). *Studies in ethnomethodology*. Englewood Cliffs: Prentice-Hall.
- Gergen, K.J. (1973). Social psychology as history. *Journal of Personality and Social Psychology*, 26(2), 309-320.
- Gergen, K.J. (1985). The social constructionist movement in modern psychology. *American Psychologist*, 40(3), 266-275.
- Gianvecchio, S., Xie, M., Wu, Z., & Wang, H. (2011). Humans and bots in internet chat: Measurement, analysis, and automated classification. *IEEE/ACM Transactions on Networking*, 19(5), 1557-1571.
- Gilbert, R.L., & Forney, A. (2015). Can avatars pass the Turing Test? Intelligent agent perception in a 3D virtual environment. *International Journal of Human-Computer Studies*, 73(1), 30-36.
- Giles, H., & Reid, S.A. (2005). Ageism across the lifespan: Towards a self-categorization model of ageing. *Journal of Social Issues*, 61(2), 389-404.
- Gillespie, A. (2007). The intersubjective dynamics of trust, distrust, and manipulation. In I. Marková & A. Gillespie (Eds.), *Trust and distrust: Sociocultural perspectives* (pp. 131-152). Greenwich: Information Age Publishing.
- Gillespie, A., Best, C., & O'Neill, B. (2012). Cognitive function and assistive technology for cognition: A systematic review. *Journal of the International Neuropsychological Society*, 18(1), 1-19.
- Gillespie, A., & Cornish, F. (2010). Intersubjectivity: Towards a dialogical analysis. *Journal for the Theory of Social Behaviour*, 40(1), 19-46.
- Gillespie, A., & Cornish, F. (2014). Sensitizing questions: A method to facilitate analyzing the meaning of an utterance. *Integrative Psychological and Behavioral Science*, 48(4), 435-452.
- Giorgi, A. (1998). The origins of the Journal of Phenomenological Psychology and some difficulties in introducing phenomenology into scientific psychology. *Journal of Phenomenological Psychology*, 29(2), 161-176.
- Glăveanu, V.P., and Lahlou, S. (2012). Through the creator's eyes: Using the subjective

- camera to study craft creativity. *Creativity Research Journal*, 24(2-3), 152-162.
- Goffman, E. (1963). *Stigma, Notes on the Management of Spoiled Identity*. London, UK: Penguin Books.
- Goldinger, S.D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105(2), 241-279.
- Goodwin, C. (2003). Introduction. In C. Goodwin (ed.), *Conversation and brain damage* (pp. 3-20). Oxford, UK: Oxford University Press.
- Goodwin, C., & Heritage, J. (1990). Conversation analysis. *Annual Review of Anthropology*, 19, 283-307.
- Gordon, D.A. (1975). A mobile, wireless “bug-in-the-ear” communication system for training and therapy. *Behavior Therapy*, 6(1), 130-132.
- Gottschalk, S. (2010). The presentation of avatars in Second Life: Self and interaction in social virtual spaces. *Symbolic Interaction*, 33(4), 501-525.
- Graziano, M.S.A. (2013). *Consciousness and the social brain*. Oxford, UK: Oxford University Press.
- Green, S. (1983). *Notes on Cyranoidia* [unpublished essay]. Stanley Milgram Papers Archive Manuscripts and Archives Department, Sterling Memorial Library, Yale University, New Haven, CT.
- Greenberg, J., & Folger, R. (1988). *Controversial issues in social research methods*. New York, NY: Springer-Verlag.
- Groom, V., Bailenson, J.N., & Nass, C. (2009). The influence of racial embodiment on racial bias in immersive virtual environments. *Social Influence*, 4(3), 231-248.
- Gruebler, A., Berenz, V., & Suzuki, K. (2011). Coaching robot behaviour using continuous physiological affective feedback. In *Proceedings of the 2011 IEEE-RAS International Conference on Humanoid Robots*, Bled, Slovenia (pp. 466-471). Piscataway: IEEE.
- Gurrin, C., Smeaton, A.F., Byrne, D., O’Hare, N., Jones, G.J.F., & O’Conner, N. (2008). *Information retrieval technology*. Lecture notes in computer science (Vol. 4993, pp. 537-542). Berlin, Germany: Springer.

- Habermas, J. (1968). *Knowledge and human interests*. Cambridge, UK: Polity Press.
- Hagg bloom, S.J., Warnick, R., Warnick, J.E., Jones, V.K., Yarbrough, G.L., ... Monte, E. (2002). The 100 most eminent psychologists of the 20th century. *Review of General Psychology*, 6(2), 139-152.
- Haney, C., Banks, C., & Zimbardo, P. (1973). Interpersonal dynamics in a simulated prison. *International Journal of Criminology and Penology*, 1, 69-97.
- Harbison, D.C., Jr., Porter, R.J., Jr., & Tobey, E.A. (1989). Shadowed and simple reaction times in stutterers and nonstutterers. *Journal of the Acoustical Society of America*, 86(4), 1277-1284.
- Harnad, S. (1991). Other bodies, other minds: A machine incarnation of an old philosophical problem. *Minds and Machines*, 1(1), 43-54.
- Harnad, S. (2000). Minds, machines and Turing: The indistinguishability of indistinguishables. *Journal of Logic, Language, and Information*, 9(4), 425-445.
- Harré, R. (2006). *Key thinkers in psychology*. London, UK: Sage Publications.
- Harwood, J., Giles, H., Clement, R., Pierson, H., & Fox, S. (1994). Perceived vitality of age categories in California and Hong Kong. *Journal of Multilingual and Multicultural Development*, 15(4), 311-318.
- Healey, E.C., & Howe, S.W. (1987). Speech shadowing characteristics of stutterers under diotic and dichotic conditions. *Journal of Communication Disorders*, 20(6), 493-506.
- Hemberg, K. (2006). *Husserl's phenomenology: Knowledge, objectivity and others*. London, UK: Continuum International Publishing Group.
- Hirji, K.F., Mehta, C.R., & Patel, N.R. (1987). Computing distributions for exact logistic regression. *Journal of the American Statistical Association*, 82(400), 1110-1117.
- Holmes, D.S. (1976). Debriefing after psychological experiments: II. Effectiveness of postdeception desensitizing. *American Psychologist*, 31(12), 868-875.
- Hoyt, C.L., Aguilar, L., Kaiser, C.R., Blascovich, J., & Lee, K. (2007). The self-protective and undermining effects of attributional ambiguity. *Journal of Experimental Social Psychology*, 43(6), 884-893.

- Hoyt, C.L., & Blascovich, J. (2007). Leadership efficacy and women leaders' responses to stereotype activation. *Group Processes and Intergroup Relations*, 10(4), 595-616.
- Husserl, E. (1931). *Cartesian meditations: An introduction to phenomenology*. Dordrecht, The Netherlands: Kluwer.
- Hutchby, I., & Wooffitt, R. (2008). *Conversation analysis*. Cambridge, UK: Polity Press.
- Icheiser, G. (1943). Structure and dynamics of interpersonal relations. *American Sociological Review*, 8(3), 302-305.
- Ikegami, E. (2011). Visualizing the networked self: Agency, reflexivity, and the social life of avatars. *Social Research*, 78(4), 1155-1184.
- Ishiguro, H. (2005). Android science: Toward a new cross-disciplinary framework. In *Proceedings of the 27th Annual Conference of the Cognitive Science Society: Toward Social Mechanisms of Android Science* (pp. 18-28).
- Ishiguro, H., & Nishio, S. (2007). Building artificial humans to understand humans. *Journal of Artificial Organs*, 10(3), 133-142.
- Jack, A.I., & Roepstorff, A. (2003). Why trust the subject? *Journal of Consciousness Studies*, 10(9-10), v-xx.
- Jackson, F. (1982). Epiphenomenal qualia. *Philosophical Quarterly*, 32(127), 127-136.
- Jacoby, S., & Ochs, E. (1995). Co-construction: An introduction. *Research on Language and Social Interaction*, 28(3), 171-183.
- Johnson M., Bradshaw, J.M., Feltovich, P.J., Jonker, C.M., van Riemsdijk, M.B., & Sierhuis, M. (2014). Coactive design: Designing support for interdependence in joint activity. *Journal of Human-Robot Interaction*, 3(1), 43-69.
- Johnson-Laird, P.N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge: Harvard University Press.
- Jones, E.E., & Nisbett, R.E. (1972). The actor and the observer: Divergent perceptions of the causes of behavior. In E.E Jones, D.E. Kanouse, H.H. Kelley, R.E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior*. Morristown: General Learning Press.

- Jonze, S. (Director/Producer/Writer), Ellison, M. (Producer), & Landay, V. (Producer). (2013). *Her* [film]. Los Angeles: Annapurna Pictures.
- Judd, C.M. (2000). Everyday data analysis in social psychology: Comparisons of linear models. In H.T. Reis, & C.M. Judd (Eds.), *Handbook of research methods in social and personality psychology*. Cambridge, UK: Cambridge University Press.
- Kahn, P.H., Jr., Ishiguro, H., Friedman, B., Kanda, T., Freier, N.G., Severson, R.L., & Miller, J. (2007). What is a human? Toward psychological benchmarks in the field of human-robot interaction. *Interaction Studies*, 8(3), 363-390.
- Kaplan, F., & Hafner, V.V. (2006). The challenges of joint attention. *Interaction Studies*, 7(2), 135-169.
- Kashy, D.A., & Kenny, D.A. (2000). The analysis of data from dyads and groups. In H.T. Reis, & C.M. Judd (Eds.), *Handbook of research methods in social and personality psychology*. Cambridge, UK: Cambridge University Press.
- Kendon, A. (1970). Movement coordination in social interaction: Some examples described. *Acta Psychologica*, 32, 101-125.
- Kennedy, A., Wilkes, A., Elder, L., & Murray, W.S. (1988). Dialogue with machines. *Cognition*, 30(1), 37-72.
- Khanna, N., & Johnson, C. (2010). Passing as black: Racial identity work among biracial Americans. *Social Psychological Quarterly*, 73(4), 380-397.
- Kidder, L.H. (1981). *Selitz, Writsman, and Cook's Research methods in social relations* (4th ed.). New York: Holt, Rinehart & Winston.
- Kiesler, S. (2005). Fostering common ground in human-robot interaction. In *Proceedings of the 14th IEEE International Workshop on Robot and Human Interactive Communication (Ro-Man 2005)*, Nashville, TN, 729-734.
- Kim, U. (1999). After the “crisis” in social psychology: The development of the transactional model of science. *Asian Journal of Social Psychology*, 2(1), 1-19.
- Kircher, T., Blümel, I., Marjoram, D., Lataster, T., Krabbendam, L., Weber, J., ... & Krach, S. (2009). Online mentalising investigated with functional MRI. *Neuroscience*

Letters, 454(3), 176-181.

- Klein, P., & Westcott, M.R. (1994). The changing character of phenomenological psychology. *Canadian Psychology*, 35(2), 133-158.
- Klimoski, R., & Mohammed, S. (1994). Team mental model: Construct or metaphor? *Journal of Management*, 20(2), 403-437.
- Komatsu, T., & Yamada, S. (2011). Adaptation gap hypothesis: how differences between users' expected and perceived agent functions affect their subjective impression. *Journal of Systemics, Cybernetics and Informatics* 9(1), 67-74.
- Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F., & Kircher, T. (2008). Can machines think? Interaction and perspective taking with robots investigated via fMRI. *PLoS One*, 3(7), e2597.
- Krämer, N.C., Bente, G., Eschenburg, F., & Troitzsch, H. (2009). Embodied conversational agents: Research prospects for social psychology and an exemplary study. *Social Psychology* 40(1), 26-36.
- Krueger, J., Heckhausen, J., & Hundertmark, J. (1995). Perceiving middle-aged adults: Effects of stereotype-congruent and incongruent information. *Journal of Gerontology*, 50B(2), 82-93.
- Kusch, M. (1995). Recluse, interlocutor, interrogator: Natural and social order in turn-of-the-century psychological research schools. *Isis*, 86(3), 419-439.
- Lafferty, J.T. (1987). *Subarctic survival situation*. Plymouth: Human Synergistics.
- Laing, R.D., Phillipson, H., & Lee, A.R. (1966). *Interpersonal perception: A theory and method of research*. London, UK: Tavistock Publications.
- Lahlou, S. (2010). Digitization and transmission of human experience. *Social Science Information*, 49(3), 291-327.
- Lahlou, S. (2011). How can we capture the subject's perspective? An evidence-based approach for the social scientist. *Social Science Information*, 50(3-4), 607-655.
- Laurens, S., & Moscovici, S. (2005). The confederate's and others' self-conversion: A neglected phenomenon. *Journal of Social Psychology*, 145(2), 191-207.

Lee, V.R. (2013). The quantified self (QS) movement and some emerging opportunities for the educational technology field. *Educational Technology*, 53(6), 39-42.

Levinas, E. (1991). *Totality and Infinity: An essay on exteriority*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Levinas, E. (1998). *Entre nous: On thinking-of-the-other*. New York: Columbia University Press.

Lewandowski, G.W., Jr., & Strohmetz, D.B. (2009). Actions can speak as loud as words: Measuring behavior in psychological science. *Social and Personality Psychology Compass*, 3(6), 992-1002.

Linell, P. (2009). *Rethinking language, mind and world dialogically*. Charlotte: Information Age Publishing.

Loebner, H. (2008). How to hold a Turin Test contest. In R. Epstein, G. Roberts, & G. Beber (Eds.), *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer* (pp. 173-179). New York: Springer.

Lubart, T.I., & Mouchiroud, C. (2004). Why does self-experimentation lead to creative ideas? [Peer commentary on “Self-experimentation as a source of new ideas: Ten examples about sleep, mood, health, and weight,” by S. Roberts]. *Behavioral and Brain Sciences*, 27(2), 269-270.

Lücking, A., & Mehler, A. (2014). On three notions of grounding of artificial dialog companions. *Science, Technology and Innovation Studies*, 10(1), 31-46.

Lunt, P. (2009). *Stanley Milgram: Understanding obedience and its implications*. Houndsills, UK: Palgrave Macmillan.

MacDorman, K.F. (2006). Introduction to the special issue on android science. *Connection Science*, 18(4), 313-317.

MacDorman, K. F., and Ishiguro, H. (2006). Toward social mechanisms of android science. *Interaction Studies* 7(2), 289-296.

MacDorman, K.F., & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies*, 7(3), 297-337.

- Marbach, E. (2007). No heterophenomenology without autophenomenology: Variations on a theme of mine. *Phenomenology and the Cognitive Sciences*, 6(1), 75-87.
- Marková, I. (2003). Constitution of the self: Intersubjectivity and dialogicality. *Culture and Psychology*, 9(3), 249-259.
- Marslen-Wilson, W.D. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*, 244, 522-523.
- Marslen-Wilson, W.D. (1985). Speech shadowing and speech comprehension. *Speech Communication*, 4(1-3), 55-73.
- Martin, A. (n. d.) *Cyranic attribution: An analysis of attributional error*. Unpublished master's thesis, City University of New York Graduate Center, New York, New York.
- Mavridis, N. (2015). A review of verbal and non-verbal human-robot interactive communication. *Robotics and Autonomous Systems*, 63(1), 22-35.
- Mavridis, N., Petychakis, M., Tsamakos, A., Toulis, P., Emami, S., Kazmi, W., ... & Tanoto, A. (2010). FaceBots: Steps towards enhanced long-term human-robot interaction by utilizing and publishing online social information. *Paladyn: Journal of Behavioral Robotics*, 1(3), 169-178.
- McCarthy, A. (2006, November 7). Cyranoid and screen performance. *Architecture, body and performance*. Lecture conducted from Brown University, Providence. Retrieved from <http://proteus.brown.edu/architecturebodyperformance/1069>.
- McCarthy, A. (2011). Responsibilities of the cyranoid citizen. In A. Sarat (Ed.), *Imagining legality: Where law meets popular culture*. Tuscaloosa: University of Alabama Press.
- McCarthy, R.A, & Warrington, E.K. (1984). A two-route model of speech production: Evidence from aphasia. *Brain*, 107(2), 463-485.
- McCarthy, R.A., & Warrington, E.K. (2001). Repeating without semantics: Surface dysphasia? *Neurocase*, 7(1), 77-87.
- McConnell, A.R., & Leibold, J.M. (2001). Relations among the implicit association test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of*

Experimental Social Psychology, 37(5), 435-442.

McLeod, P., & Posner, M.I. (1984). Privileged loops from percept to act. In H. Bouma & D. Bouwhuis (Eds.), *Attention and performance X*. Mahwah: Lawrence Erlbaum Associates.

Martin, J. (2015). *Ernest Becker and Stanley Milgram: Mid-Twentieth century students of evil*. Manuscript submitted for publication.

Maybin, J. (2001). Language, struggle, and voice: The Bakhtin/Volosinov writings. In M. Wetherell, S. Taylor, & S.J. Yates (Eds.), *Discourse theory and practice: A reader* (pp. 64-71). London, UK: Sage Publications.

Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology*, 67(4), 371-378.

Milgram, S. (1967). The small word problem. *Psychology Today*, 2(1), 60-67.

Milgram, S. (1974). *Obedience to authority: An experimental view*. New York: Harper & Row.

Milgram, S. (1977). *Cyranic speech: Pilot studies* [video prepared for National Science Foundation review panel]. Stanley Milgram Papers Archive (Box 26.1). Manuscripts and Archives Department, Sterling Memorial Library, Yale University, New Haven, CT.

Milgram, S. (1977). *The person as medium: Studies in the cyranic mode* [lab report]. Stanley Milgram Papers Archive. Manuscripts and Archives Department, Sterling Memorial Library, Yale University, New Haven, CT.

Milgram, S. (1979). *The technique of mediated speech as a tool in social psychology* [National Science Foundation grant proposal]. Stanley Milgram Papers Archive. Manuscripts and Archives Department, Sterling Memorial Library, Yale University, New Haven, CT.

Milgram, S. (1981). *Cyranoids* [videos, three in series]. Stanley Milgram Papers Archive (Box 26.11, Box 26.12, Box 26.13). Manuscripts and Archives Department, Sterling Memorial Library, Yale University, New Haven, CT.

Milgram, S. (1982). *The technique of mediated speech as a tool in social psychology* [application for a PSC-CUNY research award]. Stanley Milgram Papers Archive Manuscripts and Archives Department, Sterling Memorial Library, Yale University, New Haven, CT.

Milgram, S. (1983, October). Remarks on cyranoids [letter addressed to S. Albert]. Stanley Milgram Papers Archive Manuscripts and Archives Department, Sterling Memorial Library, Yale University, New Haven, CT.

Milgram, S. (1984, August). Cyranoids [audio recording of APA symposium speech]. Stanley Milgram Papers Archive Manuscripts and Archives Department, Sterling Memorial Library, Yale University, New Haven, CT.

Milgram, S. (1992). Cyranoids. In J. Sabini & M. Silver (Eds.), *The individual in a social world: Essays and experiments* (2nd ed.). New York: McGraw-Hill.

Milgram, S. (2010). Cyranoids. In T. Blass (Ed.), *The individual in a social world: Essays and experiments* (pp. 402–409). New York: McGraw-Hill.

Milgram, S. (2010). Introduction. In T. Blass (Ed.), *The individual in a social world: Essays and experiments* (pp. xviii-xxx). New York: McGraw-Hill.

Milgram, S. (2010). On maintaining social norms: A field experiment in the subway. In T. Blass (Ed.), *The individual in a social world: Essays and experiments* (p. 34-41). New York: McGraw-Hill.

Milgram, S. (2010). Response to intrusion into waiting lines. In T. Blass (Ed.), *The individual in a social world: Essays and experiments* (p. 42-55). New York: McGraw-Hill.

Milgram, S. (n. d.). [Untitled transcription of audio notes]. Stanley Milgram Papers Archive. Manuscripts and Archives Department, Sterling Memorial Library, Yale University, New Haven, CT.

Miller, A.G. (2013, August 8). *The obedience experiments at 50: Why are they (still) so prominent, famous and problematic?* Keynote speech presented at the 2013 Obedience to Authority Conference: Milgram's Experiments 50 Years On, Nipissing University, Bracebridge, Canada.

- Miller, A.G., Collins, B.E., & Brief, D.E. (1995). Perspectives on obedience to authority: The legacy of the Milgram experiments. *Journal of Social Issues*, 51(3), 1-19.
- Mischel, T. (1970). Wundt and the conceptual foundations of psychology. *Philosophy and Phenomenological Research*, 31(1), 1-26.
- Mitchell, R. (2009). An in your face interface: Revisiting cyranoids as revealing medium for interpersonal interaction. In I.H.C. Wouters, F.P.F. Kimman, R. Tieben, S.A.M. Offermans, & H.A.H. Nagtzaam (Eds.), *Proceedings of the 5th Student Interaction Design Research Conference: Flirting With the Future* (pp. 56-59). Eindhoven, The Netherlands: Eindhoven University of Technology.
- Mitchell, R. (2010). *Teaching via human avatar: Enlivening delivery through students acting as proxies for remote lecturers*. Paper presented at SOLSTICE 2010, Lancashire, United Kingdom.
- Mitchell, R., Gillespie, A., & O'Neill, B. (2011). *Cyranic contraptions: Using personality surrogates to explore ontologically and socially dynamic contexts*. Paper presented at DESIRE'11, Eindhoven, The Netherlands.
- Mitchell, W.J., Szerszen, K.A., Sr., Lu, A.S., Schermerhorn, P.W., Scheutz, M., & MacDorman, K.F. (2011). A mismatch in the human realism of face and voice produces an uncanny valley. *I-Perception*, 2(1), 10-12.
- Mitterer, H., & Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition*, 109(1), 168-173.
- Mizumoto, T., Nakadai, K., Yoshida, T., Takeda, R., Otsuka, T., Takahashi, T., & Okuno, H.G. (2011). Design and implementation of selectable sound separation on the Texai telepresence system using HARK. In *Proceedings of the 2011 IEEE International Conference on Robotics and Automation* (pp. 2130-2137). New York: IEEE Press.
- Modig, M. (2003). *Den nödvändiga olydnaden*. Stockholm, Sweden: Natur och Kultur.
- Moghaddam, F.M. (1987). Psychology in the three worlds: As reflected by the crisis in social psychology and the move toward indigenous third-world psychology. *American*

Psychologist, 42(10), 912-920.

Montepare, J.M., & Zebrowitz, L.A. (1998). Person perception comes of age: The salience and significance of age in social judgments. *Advances in Experimental Social Psychology*, 30, 93-161.

Mori, M. (1970). *Bukimi no tani*. *Energy*, 7, 33-35.

Moscovici, S. (1976). *Social influence and social change*. London, UK: Academic Press.

Murphrey, T. (2001). Exploring conversational shadowing. *Language Teaching Research*, 5(2), 128-155.

Murphy, K.R., Herr, B.M., Lockhart, M.C., & Maguire, E. (1986). Evaluating the performance of paper people. *Journal of Applied Psychology*, 71(4), 654-661.

Muscarella, F., & Cunningham, M.R. (1996). The evolutionary significance and social perception of male pattern baldness and facial hair. *Ethology and Sociobiology*, 17(2), 99-117.

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, 83(4), 435-450.

Nakadai, K., Takahashi, T., Okuno, H.G., Nakajima, H., Hasegawa, Y., & Tsujino, H. (2010). Design and implementation of robot audition system “HARK” – open source software for listening to three simultaneous speakers. *Advanced Robotics*, 24(5-6), 739-761.

Nakamura, L. (2002). *Cybertypes: Race, ethnicity, and identity on the internet*. New York: Routledge.

Nakanishi, H., Tanaka, K., & Wada, Y. (2014). Remote handshaking: Touch enhances video-mediated social telepresence. In M. Jones, P. Palanque, A. Schmidt, & T. Grossman (Eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2143-2152). New York: ACM.

Namy, L.L., Nygaard, L.C., & Sauerteig, D. (2002). Gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology*, 21(4), 422-432.

National Research Act, 42 U.S.C. (1974).

- Nave, C.S., Sherman, R.A., Funder, D.C., Hampson, S.E., & Goldberg, L.R. (2010). On the contextual independence of personality: Teachers' assessments predict directly observed behavior after four decades. *Social Psychological and Personality Science*, 3(1), 1-9.
- Neff, M., Wang, Y., Abbott, R., & Walker, M. (2010). Evaluating the effect of gesture and language on personality perception in conversational agents. In J. Allbeck, N. Badler, T. Bickmore, C. Pelachaud, & A. Safanova (Eds.), *Intelligent virtual agents*. Berlin, Germany: Springer.
- Neuringer, A. (1981). Self-experimentation: A call for change. *Behaviorism*, 9(1), 79-94.
- Nicholson, I. (2011). "Torture at Yale": Experimental subjects, laboratory torment and the "rehabilitation" of Milgram's "obedience to authority." *Theory and Psychology*, 21(6), 737-761.
- Nicks, S.D., Korn, J.H., & Mainieri, T. (1997). The rise and fall of deception in social psychology and personality research, 1921 to 1994. *Ethics and Behavior*, 7(1), 69-77.
- Nishio, S., Ishiguro, H., Anderson, M., & Hagita, N. (2008). Expressing individuality through teleoperated android: A case study with children. In D. Cunliffe (Ed.), *Proceedings of the 3rd IASTED International Conference on Human Computer Interaction* (p. 297-302). Anaheim: ACTA Press.
- Nishio, S., Ishiguro, H., & Hagita, N. (2007). Can a teleoperated android represent personal presence? A case study with children. *Psychologia*, 50(4), 330-342.
- Nishio, S., Ishiguro, H., & Hagita, N. (2007). Geminoid: Teleoperated android of an existing person. In A.C.D.P. Filho (Ed.), *Humanoid Robots: New Developments* (pp. 343-352). Vienna, Austria: I-Tech.
- Nitsch, V., & Popp, M. (2014). Emotions in robot psychology. *Biological Cybernetics*, 108(5), 621-629.
- Nosek, B.A., & Banaji, M.R. (2001). The go/no-go association task. *Social Cognition*, 19(6), 625-666.
- Nowak, K.L., & Rauh, C. (2005). The influence of the avatar on online perceptions of anthropomorphism, androgyny, credibility, homophily, and attraction. *Journal of*

- Computer-Mediated Communication, 11*(1), 153-178.
- Nowak, P. (2012). Deceptibots: When machines go bad. *New Scientist, 214*(2870), 45-47.
- Nunamaker, J.F., Derrick, D.C., Elkins, A.C., Burgoon, J.K., & Patton, M.W. (2011). Embodied conversational agent-based kiosk for automated interviewing. *Journal of Management Information Systems, 28*(1), 17-48.
- Ogawa, K., Taura, K., Nishio, S., & Ishiguro, H. (2012). Effect of perspective change in body ownership transfer to teleoperated android robot. In *Proceedings of the 21st IEEE International Symposium on Robot and Human Interactive Communication* (pp. 1072-1077).
- O'Neill, B., & Gillespie, A. (2008). Simulating naturalistic instruction: The case for a voice mediated interface for assistive technology for cognition. *Journal of Assistive Technologies, 2*(2), 22-31.
- Open Science Collaboration, The (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), 943.
- Oppenheimer, M. (2013). *Stanley Milgram's obedience experiments and alternative forms of representing conflict*. Paper presented at the 8th Annual International Interdisciplinary Conference: Representation, Politics and Violence, Brighton, United Kingdom.
- Orne, M.T., & Holland, C.H. (1968). On the ecological validity of laboratory deceptions. *International Journal of Psychiatry, 6*(4), 282-293.
- Pardo, J.S., Jordan, K., Mallari, R., Scanlon, C., & Lewandowski, E. (2013). Phonetic convergence in shadowed speech: The relation between acoustic and perceptual measures. *Journal of Memory and Language, 69*(3), 183-195.
- Parise, S., Kiesler, S., Sproull, L., & Waters, K. (1999). Cooperating with life-like interface agents. *Computers in Human Behavior, 15*(2), 123-142.
- Patterson, M.L. (2008). Back to social behavior: Mining the mundane. *Basic and Applied Social Psychology, 30*(2), 93-101.
- Pawlak, L. (Writer/Director/Producer), & del Paso, J. (Producer). (2012). *The inspection house (Training for the family in how to act)* [Motion picture]. United Kingdom: The

Samsonov Film/Theatre Co-Operative.

Pawlak, L. (2009). *The St. Unicorn's Trust* [Performance art]. Retrieved from <http://www.mradamjames.com/#The-St-Unicorns-Trust>.

Payr, S. (2001). The virtual other: Aspects of social interaction with synthetic characters. *Applied Artificial Intelligence, 15*(6), 493-519.

Peña, J., Hancock, J.T., & Merola, N.A. (2009). The priming effects of avatars in virtual settings. *Communication Research, 36*(6), 838-856.

Perlis, D., Purang, K., & Andersen, C. (1998). Conversational adequacy: Mistakes are the essence. *International Journal of Human-Computer Studies, 48*(5), 553-575.

Perry, G. (2012). *Behind the shock machine: The untold story of the notorious Milgram psychology experiments*. New York: The New Press.

Petty, R.E., & Cacioppo, J.T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change*. New York: Springer.

Pieraccini, R. (2012). *The voice in the machine: Building computers that understand speech*. Cambridge, MA: MIT Press.

Plumert, J. M., & Nichols-Whitehead, P. (1996). Parental scaffolding of young children's spatial communication. *Developmental Psychology, 32*(3), 523–532.

Popper, K.R. (1979). *Three Worlds*. Lecture delivered at the University of Michigan. Retrieved from: <https://www.thee-online.com/Documents/Popper-3Worlds.pdf>

Popper, K.R., & Eccles, J.C. (1977). *The self and its brain: An argument for interactionism*. New York: Springer International.

Prendinger, H., & Ishizuka, M. (2005). The empathic companion: A character-based interface that addresses users' affective states. *Applied Artificial Intelligence, 19*(3-4), 267-285.

Proudfoot, D. (2011). Anthropomorphism and AI: Turing's much misunderstood imitation game. *Artificial Intelligence, 175*(5-6), 950-957.

Rafalovich, A. (2006). Making sociology relevant: The assignment and application of breaching experiments. *Teaching Sociology, 34*(2), 156-163.

Raine, R. (2009). Making a clever intelligent agent: The theory behind the implementation. In *Proceedings of the IEEE International Conference on Intelligent Computing and Intelligent Systems* (pp. 398-402).

Ranky, G.N., & Ranky, P.G. (2005). Japanese prototype service robot R&D trends and examples. *Industrial Robot*, 32(6), 460-464.

Raudaskoski, P. (1990). Repair work in human-computer interaction: A conversation analytic perspective. In P. Luff, N. Gilbert, & D. Frohlich (Eds.), *Computers and conversation* (pp. 151-172). London, UK: Academic Press.

Raudaskoski, P., & Mitchell, R. (2013). The situated accomplishment (aesthetics) of being a cyranoid. In H. Melkas & J. Buur (Eds.), *Proceedings of the Participatory Innovation Conference* (pp. 126-129). Lappeenranta, Finland: LUT Scientific and Expertise Publications.

Reicher, S.D., Haslam, S.A., & Miller, A.G. (2014). What makes a person a perpetrator? The intellectual, moral, and methodological arguments for revisiting Milgram's research on the influence of authority. *Journal of Social Issues*, 70(3), 393-408.

Renfrow, D.G. (2004). A cartography of passing in everyday life. *Symbolic Interaction*, 27(4), 485-506.

Riek, L.D., Rabinowitch, T., Chakrabarti, B., & Robinson, P. (2009, March). How anthropomorphism affects empathy toward robots. In M. Scheutz, & F. Michaud (Eds.), *Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction* (p. 245-246). New York: ACM Press.

Reis, H.T., & Gosling, S.D. (2010). Social psychological methods outside the laboratory. In S.T. Fiske, D.T. Gilbert, & G. Lindzey (Eds.), *Handbook of Social Psychology* (5th ed., Vol. 1, pp. 82-114).

Roberts, S. (2004). Self-experimentation as a source of new ideas: Ten examples about sleep, mood, health, and weight. *Behavioral and Brain Sciences*, 27(2), 227-262.

Roberts, S. (2012). The reception of my self-experimentation. *Journal of Business Research*, 65(7), 1060-1066.

Robinson, D.K. (2001). Reaction-time experiments in Wundt's institute and beyond. In R.W.

Rieber & D.K. Robinson (Eds.), *Wilhelm Wundt in history: The making of a scientific psychology*. New York: Plenum Publishers.

Rommetveit, R. (1974). *On message structure: A framework for the study of language and communication*. London, UK: John Wiley and Sons.

Rosenthal, R. (1966). *Experimenter effects in behavioral research*. East Norwalk: Appleton-Century-Crofts.

Rosenthal-von der Pütten, A.M., & Krämer, N.C. (2014). How design characteristics of robots determine evaluation and uncanny valley related responses. *Computers in Human Behavior*, 36, 422-439.

Sacks, H., Schegloff, E.A., & Jefferson, G. (1974). *A simplest systematics for the organization of turn-taking for conversation*. *Language*, 50(4), 696-735.

Sagar, A.H., & Schofield, J.W. (1980). Racial and behavioral cues in black and white children's perceptions of ambiguously aggressive acts. *Journal of Personality and Social Psychology*, 39(4), 590-598.

Sakamoto, D., Kanda, T., Ono, T., Ishiguro, H., & Hagita, N. (2007). Android as a telecommunication medium with a human-like presence. In A. Schultz, C. Breazeal, T. Fong, & S. Kiesler (Eds.), *Proceedings of the 2nd ACM/IEEE International Conference on Human-Robot Interaction* (pp. 193-200). New York: ACM.

Saygin, A.P., Chaminade, T., Ishiguro, H., Driver, J., & Frith, C. (2012). The thing that should not be: Predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social Cognitive and Affective Neuroscience*, 7(4), 413-422.

Saygin, A.P., & Stadler, W. (2012). The role of appearance and motion in action prediction. *Psychological Research*, 76(4), 388-394.

Schegloff, E.A. (1986). The routine as achievement. *Human Studies*, 9(2-3), 111-151.

Schegloff, E.A. (1992). Repair after next turn: The last structurally provided defense of intersubjectivity in conversation. *American Journal of Sociology*, 97(5), 1295-1345.

Schegloff, E.A. (1993). Reflections on quantification in the study of conversation. *Research on Language and Social Interaction*, 26(1), 99-128.

- Schegloff, E.A. (1997). Practices and actions: Boundary cases of other-initiated repair. *Discourse Processes, 23*(3), 499-545.
- Schegloff, E.A. (2000). When 'others' initiate repair. *Applied Linguistics, 21*(2), 205-243.
- Schegloff, E.A. (2007). *Sequence organization in interaction: A primer in conversation analysis* (Vol. 1). Cambridge, UK: Cambridge University Press.
- Schegloff, E.A., Jefferson, G., & Sacks, H. (1977). *The preference for self-correction in the organization of repair in conversation*. *Language, 53*(2), 361-382.
- Schein, E.H. (1996). Kurt Lewin's change theory in the field and in the classroom: Notes toward a model of managed learning. *Systems Practice, 9*(1), 27-47.
- Schönbrodt, F.D., & Asendorpf, J.B. (2011). The challenge of constructing psychologically believable agents. *Journal of Media Psychology, 23*(2), 100-107.
- Schumacher, R.M., & Czerwinski, M.P. (1992). Mental models and the acquisition of expert knowledge. In R.R. Hoffman (Ed.), *The psychology of expertise*. New York: Springer-Verlag.
- Schumaker, R.P., Ginsburg, M., Chen, H., & Liu, Y. (2007). An evaluation of the chat and knowledge delivery components of a low-level dialog system: the AZ-ALICE experiment. *Decision Support Systems, 42*(4), 2236-2246.
- Schwitzgebel, R.K., & Taylor, R.W. (1980). Impression formation under conditions of spontaneous and shadowed speech. *Journal of Social Psychology, 110*(2), 253-263.
- Searle, J.R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3*(3), 417-424.
- Selvarajah, K., & Richards, D. (2005). The use of emotions to create believable agents in a virtual environment. In M. Pechoucek, D. Steiner, & S. Thompson (Eds.), *Proceedings of the 4th International Joint Conference on Autonomous Agents and Multiagent Systems* (pp. 13-20). New York: ACM.
- Seyama, J., & Nagayama, R.S. (2007). The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence: Teleoperators and Virtual Environments, 16*(2), 135-147.

Environments, 16(4), 337-351.

Shahri, H.H., & Perlis, D. (2008). Finding ontological correspondences for a domain-independent natural language dialog agent. In M. Goker & K. Kaigh (Eds.), *Proceedings of the 20th Innovative Applications of Artificial Intelligence Conference* (pp. 1685-1692). AAAI Press.

Shockley, K., Sabadini, L., & Fowler, C.A. (2004). Imitation in shadowing words. *Perception & Psychophysics, 66*(3), 422-429.

Shrout, P.E., & Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420-428.

Sidnell, J. (2010). *Conversation analysis: An introduction*. Chichester, UK: Wiley-Blackwell.

Silverman, I. (1971). Crisis in social psychology: The relevance of relevance. *American Psychologist, 26*(6), 583-584.

Skelt, L. (2012). Dealing with misunderstandings: The sensitivity of repair in hearing impaired conversation. In M. Egbert, & A. Deppermann (Eds.), *Hearing aids communication: Integrating social interaction, audiology and user centered design to improve communication with hearing loss and hearing technologies* (pp. 56-66). Mannheim, Germany: Verlag.

Snyder, M., & Stukas, A.A., Jr. (1999). Interpersonal processes: The interplay of cognitive, motivational, and behavioral activities in social interaction. *Annual Review of Psychology, 50*(1), 273-303.

Snyder, M., Tanke, E.D., & Berscheid, E. (1977). Social perception and interpersonal behavior. On the self-fulfilling nature of social stereotypes. *Journal of Personality and Social Psychology, 35*(9), 656-666.

Spence, C., & Read, L. (2003). Speech shadowing while driving: On the difficulty of splitting attention between eye and ear. *Psychological Science, 14*(3), 251-256.

Spexard, T.P., Hanheide, M., & Sagerer, G. (2007). Human-oriented interaction with an anthropomorphic robot. *IEEE Transactions on Robotics, 23*(5), 852-862.

Straub, I., Nishio, S., & Ishiguro, H. (2010). Incorporated identity in interaction with a

teleoperated android robot: A case study. In *Proceedings of the 19th IEEE International Symposium on Robot and Human Interactive Communication* (pp. 119-124).

Strenski, I. (1982). Malinowski: Second positivism, second romanticism. *Man*, 17(4), 766-771.

Tajfel, H., Sheikh, A.A., & Gardner, R.C. (1964). Content of stereotypes and the inference of similarity between members of stereotyped groups. *Acta Psychologica*, 22, 191-201.

Takooshian, H. (2000). How Stanley Milgram taught about obedience and social influence. In T. Blass (Ed.), *Obedience to authority: Current perspectives on the Milgram paradigm* (pp. 9-24). Mahwah: Lawrence Erlbaum Associates.

Taylor, S.E., & Fiske, S.T. (1975). Point of view and perceptions of causality. *Journal of Personality and Social Psychology*, 32(3), 439-445.

ten Have, P. (1999). *Doing conversation analysis: A practical guide*. London, UK: Sage Publications.

Tesch, F.E. (1977). Debriefing research participants: Though this be method there is madness to it. *Journal of Personality and Social Psychology*, 35(4), 217-224.

Tinwell, A., Grimshaw, M., & Williams, A. (2011). Uncanny speech. In M. Grimshaw (Ed.), *Game sound technology and player interaction: Concepts and developments* (pp. 213-234). Hershey: Information Science Reference.

Tirassa, M., & Bosco, F.M. (2008). On the nature and role of intersubjectivity in human communication. In F. Morganati, A. Carassa, & G. Riva (Eds.), *Enacting intersubjectivity: A cognitive and social perspective on the study of interactions* (pp. 81-95). Amsterdam, The Netherlands: IOS Press.

Titchener, E.B. (1927). *Experimental psychology: A manual of laboratory of practice* (Vol. 1). New York: Macmillan.

Tomasello, M. (2008). *Origins of Human Communication*. Cambridge, MA: MIT Press.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*,

28(5), 721-727.

Travers, J., & Milgram, S. (1969). An experimental study of the small world problem. *Sociometry*, 32(4), 425-443.

Trevarthen, C., & Aitken, K.J. (2001). Infant intersubjectivity: Research, theory, and clinical applications. *Journal of Child Psychology and Psychiatry*, 42(1), 3-48.

Turkle, S. (2007). Authenticity in the age of digital companions. *Interaction Studies*, 8(3), 501-517.

Turing, A.M. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433-460.

UCLA Statistical Consulting Group. (2014). *Stata data analysis examples: Exact logistic regression*. Retrieved from <http://www.ats.ucla.edu/stat/stata/dae/exlogit.htm>.

Varela, F.J., & Shear, J. (1999). First-person methodologies: What, why, how? *Journal of Consciousness Studies*, 6(2-3), 1-14.

Vasalou, A., & Joinson, A.N. (2009). Me, myself and I: The role of interactional context on self-presentation through avatars. *Computers in Human Behavior*, 25(2), 510-520.

Vermersch, P. (1999). Introspection as practice. *Journal of Consciousness Studies*, 6(2-3), 17-42.

Visser, T., Traum, D., DeVault, D., & op den Akker, R. (2014). A model for incremental grounding in spoken dialogue systems. *Journal of Multimodal User Interfaces*, 8(1), 61-73.

Vogeley, K., & Bente, G. (2010). “Artificial humans”: Psychology and neuroscience perspectives on embodiment and nonverbal communication. *Neural Networks*, 23(8-9), 1077-1090.

Vygotsky, L. (1978). Mind in society: The development of higher psychological processes. Cambridge: Harvard University Press.

Wagoner, B. (2009). The experimental methodology of constructive microgenesis. In J. Valsiner, P.C.M. Molenaar, M.C.D.P. Lyra, & N. Chaudhary (Eds.), *Dynamic process methodology in the social and developmental sciences*. New York: Springer.

- Wallace, R. (2008). The anatomy of A.L.I.C.E. In R. Epstein, G. Roberts, & G. Beber (Eds.), *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer* (pp. 181-210). New York: Springer.
- Wallace, R. (2015). A.L.I.C.E. [computer program]. Retrieved from <http://alice.pandorabots.com>.
- Watanabe, M., Ogawa, K., & Ishiguro, H. (2014). Field study: Can androids be a social entity in the real world? In G. Sagerer, M. Imai, T. Belpaeme, & A. Thomaz (Eds.), *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 316-317). New York: ACM.
- Watt, S. (2008). Can people think? Or machines? In R. Epstein, G. Roberts, & G. Beber (Eds.), *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer* (pp. 301-318). New York: Springer.
- Watts, D.J., & Strogatz, S.H. (1998). Collective dynamics of “small-world” networks. *Nature*, 393, 440-442.
- Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, 52, 113-117.
- Weizenbaum, J. (1966). Eliza – a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36-45.
- Wheeler, S.C., & Petty, R.E. (2001). The effects of stereotype activation on behaviour: A review of possible mechanisms. *Psychological Bulletin*, 127(6), 797-826.
- Wilcox, B. (2015). Rose [computer program]. Retrieved from <http://brilligunderstanding.com/rosedemo.html>.
- Willard, J., Madon, S., Guyll, M., Scherr, K.C., & Buller, A.A. (2012). The accumulating effects of shared expectations. *European Journal of Social Psychology*, 42(4), 497-508.
- Williams, M. (2007). Avatar watching: Participant observation in graphical online environments. *Qualitative Research*, 7(1), 5-24.

- Wilson, T.D. (2003). Knowing when to ask: Introspection and the adaptive unconscious. *Journal of Consciousness Studies, 10*(9-10), 131-140.
- Winston, A.S., & Blais, D.J. (1996). What counts as an experiment? A transdisciplinary analysis of textbooks, 1930-1970. *American Journal of Psychology, 109*(4), 599-616.
- Worswick, S. (2015). *Mitsuku* [computer program]. Retrieved from:
<http://www.square-bear.co.uk/mitsuku/housebot.htm>.
- Yee, N., & Bailenson, J.N. (2006). *Walk a mile in digital shoes: The impact of embodied perspective-taking on the reduction of negative stereotyping in immersive virtual environments*. Paper presented at PRESENCE 2006: The 9th Annual International Workshop on Presence, Cleveland.
- Yee, N., & Bailenson, J.N. (2007). The Proteus effect: The effect of transformed self-representation on behavior. *Human Communication Research, 33*(3), 271-290.
- Yee, N., & Bailenson, J.N. (2009). The difference between being and seeing: The relative contribution of self-perception and priming to behavioral changes via digital self-representation. *Media Psychology, 12*(2), 195-209.
- Yee, N., & Bailenson, J.N., & Ducheneaut, N. (2009). The Proteus effect: Implications of transformed digital self-representation on online and offline behavior. *Communication Research, 36*(2), 285-312.
- Yong, E. (2012). Bad copy: In the wake of high-profile controversies, psychologists are facing up to problems with replication. *Nature, 485*, 298-300.
- Yong, E. (2012). Nobel laureate challenges psychologists to clean up their act: Social-priming research needs “daisy chain” of replication. *Nature*. Retrieved from <http://www.nature.com/news/nobel-laureate-challenges-psychologists-to-clean-up-their-act-1.11535>.
- You, J. (2015). Beyond the Turing Test. *Science, 347*(6218), 116.
- Young, G., & Whitty, M.T. (2010). In search of the Cartesian self: An examination of disembodiment within 21st-century communication. *Theory & Psychology, 20*(2), 209-229.

- Zahn, C.J. (1984). A reexamination of conversational repair. *Communication Monographs*, 51(1), 56-66.
- Zdenek, S. (2001). Passing Loebner's Turing Test: A case of conflicting discourse functions. *Minds and Machines*, 11(1), 53-76.
- Zebrowitz, L.A. (2002). The affordances of immersive virtual environment technology for studying social affordances. *Psychological Inquiry*, 13(2), 143-145.
- Ziemke, T., & Lindblom, J. (2006). Some methodological issues in android science. *Interaction Studies*, 7(3), 339-342.
- Zimbardo, P. (1992). [Foreword]. In. J. Sabini & M. Silver (Eds.), *The individual in a social world: Essays and experiments* (2nd ed., pp. ix-xi). New York: McGraw-Hill.
- Zimbardo, P. (1999). Experimental social psychology: Behaviorism with minds and matters. In A. Rodrigues & R.V. Levine (Eds.), *Reflections on 100 years of experimental social psychology* (pp. 135-157), New York: Basic Books.
- Zue, V.W., & Glass, J.R. (2000). Conversational interfaces. Advances and challenges. *Proceedings of the IEEE*, 88(8), 1166-1180.

APPENDICES | A: Experiment Materials from Chapter 2

Study 1: Participant instructions

Instructions for participants

This study is designed to explore how strangers talk to one another.

In a few moments, you will be introduced to another research participant. You are asked to discuss a topic (or topics) of *your choosing* with the other research participant for 10 minutes.

Feel free to talk with the other participant about whatever you like, but please do not ask any questions that would divulge close personal information (e.g. family details) or that would be inappropriate or vulgar.

After the discussion is over, you will fill out a questionnaire regarding how the discussion went.

The researcher will not be present in the room during the discussion, so if you have any questions regarding these instructions, please inform the researcher now.

Before you enter the discussion room, please take a few moments to choose a topic (or topics) you would like to discuss with your discussion partner (however, do not disclose this topic with the researcher until *after* your discussion).

Study 2: Participant instructions

Participant Instructions

Participant,

In a few moments you will be introduced to an individual. I (the researcher) am interested in knowing about the intellectual potential of this person.

Rather than assess this myself, I am asking research participants (YOU) to act as independent interviewers.

You will have **20 minutes** to interview the individual. Your task is to simply question the individual to get a good picture of what they're like and what they know. **Ask whatever questions you feel will best reveal the person's intellectual potential, but your questions should touch upon these areas:**

- Science
- Literature
- Current or Historic Political Events

Try to probe the limits of the person's knowledge in these areas.

The interviewee has not been provided with answers and will not be using a script.

As time is limited, please do not ask questions or make comments that are unrelated to addressing the person's intellectual potential.

You may not discuss or ask questions related to the following issues:

- The person's age or closely related personal details (e.g. when they were born).
- The person's educational history (e.g. where they attend or have attended school).

** The person you will interview has been instructed not to answer questions that would reveal this and closely related information*

During the interview, you may not talk directly to the other participant interviewers. Please direct all questions/comments to the person you are interviewing. Feel free to expand upon a point made by another interviewer, or to continue a line of questioning initiated by another interviewer, but refrain from issuing questions/comments to the other interviewers.

The researcher will not be present in the room during your discussion with the interviewee, so if you have any questions about the study or need any instructions clarified, please alert the researcher now.

You may use the front or reverse side of this sheet to take notes.

Debriefing Form (administered to participants who were subjected to deception)

DEBRIEFING FORM

Now that you have discussed with the researchers the full nature of this study, you are now being asked to indicate your attitudes toward research of this nature. Should you choose to withdraw your data from this study, you are now being given the opportunity to do so. The researcher will read through this form aloud with you to help you understand its contents. **Please sign the bottom of this form indicating that you AGREE with all of the following statements.**

- I (research participant) have been debriefed by the researchers as to the full nature of the research study I have taken part in.
- I understand that the nature of this study involved me speaking with an individual who was channeling the words of another person.
- I would not object to another individual such as myself taking part in this study, should they consent to do so.
- I consent to the researchers using all data collected from me.
- I am aware of my right to anonymity with regards to this study and understand that the data I have given will be kept confidential.

Print Name: _____

Date: _____ / _____ / _____

Your Signature: _____

APPENDICES | B: Experiment Materials from Chapter 3

Study 1: Participant instructions (text interface condition)

Con: CO

INSTRUCTIONS FOR PARTICIPANTS

You are being asked to engage in a 10-minute conversation with 2 different interlocutors: **Interlocutor A** and **Interlocutor B**.

One of these interlocutors is a human being. The other is merely a computer program known as a “chat bot” that is designed to imitate human communication. The chat bot has been given a human name, and will respond as if it were a human being.

After the 10-minute conversation ends, you will be asked to decide which interlocutor is the human, and which is the computer program.

You may discuss whatever you’d like with the interlocutors, but please refrain from being vulgar.

--

You will use the computer interface in front of you to communicate with the interlocutors. Please type your comments using the keyboard provided.

Use the text window on the left to communicate with Interlocutor A, and the text window on the right to communicate with Interlocutor B. Please do not close either window.

Please be aware that Interlocutor A and Interlocutor B *are not in conversation with one another*. You are the only person in communication with either of them. Any message you send to Interlocutor A will only be received by Interlocutor A (and vice versa).

--

Your identity and all of your communications will be kept strictly confidential. Furthermore, all data you provide will be completely anonymized. You may withdraw from this study at any time for any reason.

Study 1: Participant instructions (echoborg condition)

Con: CONFE

INSTRUCTIONS FOR PARTICIPANTS

You are being asked to engage in a 10-minute conversation with 2 different interlocutors: **Interlocutor A** and **Interlocutor B**.

One of these interlocutors is a human being. The other is merely a computer program known as a “chabot” that is designed to imitate human communication. The chatbot has been given a human name, and will respond as if it were a human being.

After the 10-minute conversation ends, you will be asked to decide which interlocutor is the human, and which is the computer program.

You may discuss whatever you’d like with the interlocutors, but please refrain from being vulgar.

You will use the computer interface in front of you to communicate with the interlocutors. Please type your comments using the keyboard provided.

Use the text window on the left to communicate with Interlocutor A, and the text window on the right to communicate with Interlocutor B. Please do not close either window.

The individuals in front of you are wearing inner ear devices that receive the responses generated by your interlocutors. The individual on the left will speak aloud to you Interlocutor A’s responses, while the individual on the right will speak Interlocutor B’s responses. The individuals in front of you will at no point speak any of their own thoughts; they are merely repeating the responses they receive via their earpieces. Your interlocutors will only respond to what you type into the computer terminal. They will not respond to anything you speak aloud. You will not see your interlocutors’ responses appear on the computer screen; they will only be spoken aloud by the individuals in front of you.

Please be aware that Interlocutor A and Interlocutor B *are not in conversation with one another*. You are the only person in communication with either of them. Any message you send to Interlocutor A will only be received by Interlocutor A (and vice versa).

--

Your identity and all of your communications will be kept strictly confidential. Furthermore, all data you provide will be completely anonymized.

You are being asked to engage in a 10-minute conversation with an Interlocutor.

Please respond to the messages that appear in the computer terminal in any manner you'd like, but please refrain from saying anything vulgar.

To send a reply, enter your response into the text terminal and press "ENTER".

Please only send 1 reply for each message you receive.

Also, please be advised that some time might elapse in between the messages you receive. This is completely normal. The conversation will last the full 10 minutes.

--

When the conversation is over, the researcher will return to the room to debrief you.

Your identity and all of your communications will be kept strictly confidential. Furthermore, all data you provide will be completely anonymized.

Study 2: Participant instructions (text interface condition)

COE

Participant,

In a moment you will engage in a 10-minute conversation with an interlocutor.

Your interaction will be recorded, but only the researcher will have access to this data, and your identity will remain confidential. This study is not designed to make you feel uncomfortable, but if you for whatever reason you feel like ending the study prior to its completion, you may do so.

Throughout the interaction, one of two scenarios will be the case:

Scenario A	Scenario B
<p>Your interlocutor will be communicating entirely the words and phrases generated by a chatbot (a computer program designed to engage in human-like conversation) and none of their own thoughts. The computer has been programmed with the identity of a human, and will respond to you as if it were a human.</p> <p>100% words/phrases from a computer program</p>	<p>Your interlocutor will be <i>imitating</i> a chatbot computer program that has the identity of a human. They will not communicate any actual words generated by a chatbot, but have been trained to communicate as though they were one.</p> <p>0% words/phrases from a computer program (Human imitating a computer)</p>

You will communicate with your interlocutor via a computer interface. Your interlocutor will not see you physically, but will respond to the words you speak.

Your interlocutor, meanwhile, will communicate with you via text. You will not see or hear your interlocutor, but will be able to read their communication on this computer screen. Please do not touch the screen or keyboard during your interaction and refrain from making any changes to the computer.

Following the interaction, the researcher will re-enter the room and you will be asked which of the two scenarios detailed above you believe to be the case.

Study 2: Participant instructions (echoborg condition)

CYE

Participant,

In a moment you will engage in a 10-minute conversation with an interlocutor.

Your interaction will be recorded, but only the researcher will have access to this data, and your identity will remain confidential. This study is not designed to make you feel uncomfortable, but if you for whatever reason you feel like ending the study prior to its completion, you may do so.

Throughout the interaction, one of two scenarios will be the case:

Scenario A	Scenario B
<p>Your interlocutor will be communicating entirely the words and phrases generated by a chatbot (a computer program designed to engage in human-like conversation) and none of their own thoughts. The computer has been programmed with the identity of a human, and will respond to you as if it were a human.</p> <p>100% words/phrases from a computer program</p>	<p>Your interlocutor will be <i>imitating</i> a chatbot computer program that has the identity of a human. They will not communicate any actual words generated by a chatbot, but have been trained to communicate as though they were one.</p> <p>0% words/phrases from a computer program (Human imitating a computer)</p>

Your interlocutor will sit in the chair opposite you.

Following the interaction, the researcher will re-enter the room and you will be asked which of the two scenarios detailed above you believe to be the case.

Study 3: Participant instructions (text interface condition)

Instructions

This study concerns how people converse when they meet each other for the first time.

In a few minutes, you will be introduced to another participant and asked to speak to them for 10 minutes. You will speak aloud toward the computer monitor, while your interlocutor will respond via text visible on the screen. There is no need to type anything.

You may speak about whatever you like during the interaction.

This study is not designed to make you feel uncomfortable, but if for whatever reason you would like to end the study before its completion, you may do so. Simply tell the researcher.

Audio will be recorded by the researcher, but no one else will have access to this data. Your participation in this study will be kept confidential.

Study 3: Participant instructions (echoborg condition)

Instructions

This study concerns how people converse when they meet each other for the first time.

In a few minutes, you will be introduced to another participant and asked to speak to them for 10 minutes.

You may speak about whatever you like during the interaction.

This study is not designed to make you feel uncomfortable, but if for whatever reason you would like to end the study before its completion, you may do so. Simply tell the researcher.

Audio will be recorded by the researcher, but no one else will have access to this data. Your participation in this study will be kept confidential.

APPENDICES | C: Experiment Materials from Chapter 4

Participant instructions (text interface condition)

co

Participant,

This study is concerned with how strangers converse with one another.

In a moment you will engage in a 10-minute conversation with an interlocutor. You may speak about whatever topics you choose, but please refrain from discussing anything vulgar. You are not required to speak about anything you feel uncomfortable speaking about.

Your interaction will be recorded, but only the researcher will have access to this data, and your identity will remain confidential.

This study is not designed to make you feel uncomfortable, but if you for whatever reason you feel like ending the study prior to its completion, you may do so.

—

You will communicate with your interlocutor via a computer interface. Your interlocutor will not see you physically, but will respond to the words you speak.

Your interlocutor, meanwhile, will communicate with you via text. You will not see or hear your interlocutor, but will be able to read their communication on this computer screen. Please do not touch the screen or keyboard during your interaction and refrain from making any changes to the computer.

After 10 minutes, the researcher will enter the room and ask you questions regarding your impressions of the conversation.

Participant instructions (echoborg condition)

CY

Participant,

This study is concerned with how strangers converse with one another.

In a moment you will engage in a 10-minute conversation with an interlocutor. You may speak about whatever topics you choose, but please refrain from discussing anything vulgar. You are not required to speak about anything you feel uncomfortable speaking about.

Your interaction will be recorded, but only the researcher will have access to this data, and your identity will remain confidential.

This study is not designed to make you feel uncomfortable, but if you for whatever reason you feel like ending the study prior to its completion, you may do so.

—

Your interlocutor will sit in the chair opposite you.

After 10 minutes, the researcher will enter the room and ask you questions regarding your impressions of the conversation.

APPENDICES | D: Example Transcripts from Chapter 2

Study 2: Adult source, child shadower (trial 1)

C: Cyranoid

I1: Female (American accent) - middle

I2: Female (American accent) - far left side.

I3: Female (American accent) - middle

I4: Female (Foreign accent) - middle

I5: Male - right

<Begin Transcript>

Panel: <Laughs>

I1: Hi! How are you doing?

C: I'm doing well.

I1: Great! we have some questions for you.

Panel: <Laughs>

I1: What's your favorite book that you've ever read?

C: Um, my favorite book, um, are plays included?

I1: Sure.

C: Ya. Cause then I would say King Lear.

I1: Ya, what do you like about it?

C: I like, um, the diversity characters. Ya.

I1: Who is your favorite character?

C: Um, King Lear.

I1: Ya?

C: Ya.

I1: Awesome.

I2: Do you often prefer plays to books?

C: No, um I like some books as well. I like, um, Dostovesky.

I2: Okay, what is your favorite Dostovesky novel?

C: Um, "The Idiot".

I2: Okay, why?

C: Because, um, it's a book about someone who is, um, like Jesus. And, in the modern world he comes across as an idiot because he's too nice. Ya.

I3: What about poetry? Do you read any of that?

C: I don't really read that much poetry.

I3: Do you, why?

C: Um, I'm not sure, but there's a lot to read. I can't read everything, so <Laughs>

I4: And when you read something like Dostovesky are you interested in, like, the historical background such, cause I know, like, his life history, about Russia and about like what was

going on at the time... Are you interested in knowing more about that?

C: Um, it's very interesting to see Russia before the revolution, so the historical context, but I mean, interesting with Dostovesky, um, from a psychological aspect, ya.

I1: So if you could conduct a science experiment on anything you wanted, what would it be?

C: Um, do I have to obey the laws of physics?

I1: <Laughs> I'll say yes.

C: Okay, um, do I have an unlimited budget? Or, ya? Then I would like to test the speed of light.

I1: <Laughs> How would you do it?

C: I would um, use my unlimited budget to build a rocket which went very fast and I would put a clock on it. And, I would have it go around the world, ya, then maybe build another rocket which went in the other direction with another clock on it, and I would, um, use my unlimited budget to make it go very very fast. Then I would see if the clocks were the same.

I4: And you would like, do some calculus, or would <inaudible>?

C: Well, I would see, if they were atomic clocks then they should be identical. But according to the theory of relativity, they shouldn't be.

I2: So do you ever read the newspaper or watch the news?

C: I don't have a television.

I2: Okay <laughs>

C: But I do read a lot of newspapers.

I2: Okay.

C: I do read blogs.

I2: Okay. So are there any current events that you've read about on your blogs lately?

C: Um, I was interested in reading, there was a 3D printer which was, um, printing, um guns.

Ya. So then now there is, um, someone uploaded, um, the diagram with schematics for printing a gun using a 3D printer. So..

I2: And what do you think about that? Is that a good thing is that a bad thing?

C: I think, it's probably a good thing.

I2: Why?

C: Because, um, knowledge is power. And people will print more than guns. So, they will print utensils, they will print jewelery, they will print, um, tables and chairs. They will print, um beautiful things. Not just guns.

I3: Nice.

I5: Um, do you follow politics?

C: A little bit.

I4: Do you know, for instance, um, the main political parties? What's going on, and stuff like that?

C: In what country?

I4: England.

C: So there's the Tories, Labour, and the Liberal Democrats, I think? Ya.. <laughs> I think

so.

Panel: <Laughs>

C: There's also the Green Party.

I4: And what would be, if you had to name a very important, um, historical event for you, that has marked you, what would you say?

C: Um, historical event...

I4: I mean, it can be very recent, or...

C: Um, do you mean marked me personally or marked history?

I4: You personally, like, what do you think was very interesting, or...

C: I find World War II very interesting.

I4: Ya? Why?

C: Because, um, a lot of clever people were very stupid.

I2: What did you find the most stupid about World War II?

C: Um, the way in which we created a situation which we couldn't escape.

I2: Are you referring to, like, appeasement, or what are you referring to?

C: I'm referring to, um, that, I guess the, um the appeasement is part of it, because that created a situations for the Russians, er, for the Germans I mean, and they felt they had to, erm, um... they wanted to get some sort of counterbalance, ya. So, like, Hitler was an answer to the appeasement of World War I. And, that was our mistake, for creating a situation in which Hitler was logical. But equally, um, the Nazi Party created a situation

for the United Kingdom and America in which war was inevitable.

I3: So you think World War II was inevitable?

C: I think we created a situation which made it inevitable. It was inevitable because it was our situation. We could have created a different situation, ya.

I2: So do you prefer international political or domestic politics?

C: International.

I2: Why?

C: Because it is, um, more significant. Ya.

I2: Are you interested in the work of any international organisations, such as the United Nations? Or are you interested in any countries foreign policies?

C: At the moment I'm interested in North Korea's foreign policy, ya.

I2: How do you think that situation is going? Or do you think it's going to stay fairly stable?

C: Um, it depends. I think, again, like, linking back to the World War II discussion, we are, um, possibly creating a situation in which North Korea will do something stupid.

I3: But do you think that there is any credibility to their threats?

C: They do have nuclear weapons. So I wouldn't stand up publicly and say they had no credibility.

I3: Mmhmm..

C: Because that might make them use their weapons.

Pane: <Laughs>

I1: What other things do you think might deter North Korea from using nuclear weapons against other countries?

C: Um, I'm not sure. That's a difficult, that's a difficult question. There's a lot of people who are trying to work that out at the moment, so. So maybe being nice to them.

Panel: <Laughs>

I3: Isn't that appeasement? Isn't that what you said created the situation in World War II?

C: Ya. If I knew the answers to these questions I'd be a politician.

Panel: <Laughs>

I3: A very dumb politician.

Panel: <Laughs>

C: It's important to recognize the limits of knowledge. Ya.

I1: So which country outside of England do you know the most about as far as politics?

C: Everybody's subjected to American politics. So...

I1: And what do you think about politics in America?

C: Um, I think it is very, um, influenced by business.

I1: Why is that? Why does business have a big influence there?

C: I read a blog recently which said there were 50, um, lobbyists for every senator.

I4: And when you say that the history and politics that you're more aware of is America, would you say that it's also relating to the American path and American culture? Do you read for instance, American literature? Things like that? Are you interested in things like that as well?

C: Well, um, we get exposed to a lot of American politics because of it's in English. And we don't get Russian politics in the same way. I'd like to know about Chinese politics, but I don't speak Chinese.

Panel: <Laughs>.

I4: Of course. But I mean, like, do you also <inaudible>, like, do you read American literature, do you see American movies, or other stuff?

C: I read um, a biography of JFK. Um, ya.

I4: And, in terms of literature, you prefer English ones or have a preference for classics, I don't know....?

C: I wouldn't make a preference based on nationality.

I4: So what is your criteria to read a book?

C: Fiction or non-fiction? For fiction, I like, um, I like diversity of characters. So, um, I'm influenced by Bakhtin. Um, ya. That's why I like Dostoevsky.

I2: Do you think literature has a role to play in how we think about history or how we think about current political events?

C: I do. But I think it has a more important role in making us think about the future. Like, literature is about alternative possibility. Ya.

I2: Now, you were obviously reading the English translation of Dostoevsky, but would you ever hope to at some point to be able to pick up a language and read the sort of classical

literature in that language.

C: Um, I'm hopeful for Google Translate <laughs>. Um, I think in the future, um languages won't exist as we know them. I think we will get real-time translation and it will be very good. So, I wouldn't put a lot of time into learning Russian.

I4: You don't think that reading in the proper language brings something additional?

C: I'm sure it does, but if was to learn it... Russian... that would take a lot of time. And there's lots think that I couldn't read because I'd be learning Russian, so...

I2: So in this hypothetical world where you have real-time translations of all languages, how do you think that would affect, I don't know, diplomacy, or, politics, or the way we do finance?

C: I think communication is good. So, um, the, um, the Cuban Missile Crisis, the White House and the Kremlin set up the Red Phone here that prevented future um, sort of misunderstandings. So I believe in communication. I believe in translation.

I4: What do you think have been the most amazing inventions for the next years?

C: Future inventions?

I4: Ya.

C: Ya. Um, if i knew them I'd be a great inventor, so <laughs>.

I4: Ya, but I mean, what are you hopes, or what would you like to invent if you had to?

C: What would I like to invent...That's a good quetsion. Um, I'd like to invent a cheap source of energy.

I4: Why?

C: That doesn't have a sort of environmental impact. Ya. I think that's the main problem we face.

I1: So you mention that you know some stuff about American politics. So, about the 2012 Presidential election, do you know who the candidates were and the differences between them?

C: Ya I think, um, I think was it Barack Obama was elected?

I1: Do you know who ran against him?

C: Um, Uh, Barack Obama he was in his second term, and I think it's good that he got elected because he, um, he could finish a project he started, but I'm worried that he's going to create a lot of debt for the United States.

I1: Do you think that the other candidate Romney would do a better job?

C: Romney is a businessman. And I would be worried that he would, um, support his friends. Ya.

I1: So concerns either way?

C: Ya <laughs>.

I1: <Laughs> Fair enough.

I2: You said you read blogs. What blogs do you read?

C: I read, um, Boing Boing.

I2: Okay.

C: The Financial Times. Singularity. The BBC. Ya.

I2: There are a lot of blogs on the internet, so the fact that you have a very few specific ones that you read time and time again, um, what draws you to these particular blogs, is it the style of their writing, is it about what they cover?

C: Well, you asked me what blogs I sort of read regularly. And those are the ones I read regularly. But I read quite a lot of other ones still.

I2: Okay, but what draws you to them? Why do you read them regularly? What do you find good about them that maybe other blogs don't necessarily do all the time?

C: Well, what I'm trying to say is that I don't rely on them too much, I, um, I like Google News where news stories covered there are links to many different articles to the same topics. And I like to read the different points of view on that topic. Ya.

I3: When you're reading things on blogs, because you're reading so many different kinds, how do you decide what, like what opinion is correct, or what side you agree with.

C: I don't think opinions are ever correct. I could see like, um, following up with references, ya. The internet is great for research. Ya.

I4: And do you use, like, social networks to conduct your news reading. I don't know, like Twitter.

C: I don't use social networks.

I4: Okay.

C: My friends do, but...

I4: Why don't you?

C: Um, I'm busy.

I4: Okay.

I2: What was the last book you read? <END>

Study 2: Child source, adult shadower (trial 3)

C: Cyranoid

I1: Male (Asian accent) mid right

I2: Female (French accent) mid left

I3: Female (Asian accent) far right

I4: Female (Asian accent) far left

<Begin Transcript>

C: Hi.

Panel: Hi.

I1: Um, maybe I can start first. Uh, I want to know what's your view on the legacy of Margaret Thatcher.

C: Very mixed, because she done some good things and some bad things, um, uh, lot's of people don't like her. Some of the good things she's done... she got the Falklands back from Argentina. Uh, and some of the bad things... She cut benefits, she, she tried to get rid of people, like foreigners. Uh, ya. Mixed, very mixed.

I1: Thank you.

I2: Uh, maybe to continue with this. What did you think about the whole controversy about her? Do you know what happened here?

C: Um, the controversy about her funeral?

I2: Ya. Like, people say, like, spending too much money on this, like state money.

C: Well, if she's already died, then I don't think, they shouldn't spend more than they have to. Because, if they spend more than they have to, then, like, it's unfair. There's loads of other people who die every week, every day even, and they don't get, like, all this, all this, ya, money spent on them. It could go to good causes, ya.

I3: So regarding her funeral, do you think that Churchill had a bigger <inaudible>

C: Sorry could you repeat that?

I3: Churchill.

C: Churchill?

I3: Ya, Churchill had even a bigger funeral than Margaret Thatcher. Do you think he deserved it better than Margaret Thatcher?

C: Um, I think yes.

I3: Why?

C: Because he, he won a victory for us. He was really big. And, like, um, he saved many people around the world. And, um, ya. If it wasn't for him, I don't think we would have won the war at all.

I3: Margaret Thatcher also helped the British to win a victory, so, compared to Churchill...

C: Ya, erm. I don't think it's, ya, I don't think it was as big as what Churchill did.

I3: <laughs>

C: The World War was much, much bigger than the Falklands War.

I3: Ya. Mmhmm.

C: It was, so, so I think Churchill should get a bit more credit than Thatcher.

I1: Do you know what, what is the freezing point of water? I mean, how hot?

C: The freezing point of water?

I1: Yes.

C: Um, is it, um, zero? zero degrees?

I1: You think it's zero degrees?

C: The freezing point?

I3: Ya, the freezing point.

C: Zero degrees celsius. I'm not sure, ya?

I3: It is, lower, it's like, minus twenty.

C: Okay.

I3: Um, if one thing you could change in the current NHS system, what would you recommend?

C: Um, do you mean about, like...

I3: Their policies...

C: For England? For the NHS?

I3: Ya.

C: Um, I'd change the NHS system, make it more organized. It's very scrappy.

I3: In what perspective?

C: In the perspective like, when you go to your doctors and you see like the cues, like, the

cues are big, and like, it's not organized. And, it's, if you want to book an appointment, it takes you hours to, to book one.

I3: So it's not efficient?

C: No. Not at all.

I3: So, um, if you think it's not efficient, what kind of recommendation you can give to change this?

C: Ya, getting advice from the public. Ya. They're the main people who they're helping, so, they should get advice from the public and see what they want. How would they change it?

I2: Um, what do you think that, I mean, the Olympics in London has been promoted as something that was going to help East London. Do you think it's going to benefit most of the people living there.

C: Yes. Ya, because, um, it's uh, it's brought something to London. It's brought people to London, um, and it's sort of a really good thing. And I think lot's of people are talking about it, talking about it, um, some people said it was over spent, that they spent too much, but I think it's good what they've done, overall. And they're building house, so it's not going to waste.

I1: So your view on the Olympic games is more on the economic side, because you talk about the spending, and it's worth it. But do you think there's other, do you have other perspectives on this? For example, like, maybe like, the Olympic games improved the social cohesion in some sort of other perspective than the economic one.

C: Um, I think, it brought, it better, like, better perspective on London. It sort of, mmm... it made the people in London feel proud. Ya. That's what I think.

I3: What do you think is the greatest innovation in this century?

C: I would say electricity.

I3: The electricity. What kind of electricity?

C: All forms of electricity. Um, if it wasn't for electricity, we wouldn't have the things we take for granted today, like t.v., uh, computers, phones, all those things.

I3: Mmhmm. So do you know what year they started to have electricity?

C: Uh, which year we started to have electricity? Um,

I3: Ya.

C: I'm not so sure exactly.

I2: Um, and the person who is supposed to have invented electricity?

C: I don't know <laughs>.

I2: <laughs> It's really different, but, what do you think about, like, they are adapting the book, um, "Gatsby"?

C: Uh, "The Great Gatsby"?

I2: Yes.

C: Um, I don't know that much about "The Great Gatsby", but, ya, I know it's in the cinema. Ya. Leonardo DiCaprio in it, maybe?

I2: But, you don't have an idea about, like, transposing this kind of novel in the cinema?

C: Ya, I think it's a good idea. It depends on what kind of people they're reaching out to. But, if they're reaching out to people who have, sort of, never read the book before, then they might find it a bit too complicated. But if they're reaching out to people who have read the book, then, they might love it.

I1: I want to talk about financial crisis. People say the bankers of the banks have created, what they call, like, systematic risk, that effects, uh, the real economy in crisis time. And then, as far as we can see, in the real world, bankers don't take a lot of responsibility. They usually have a very huge amount of bonus even if the bank goes nearly bankrupt. So, I want to know what's your view on this? Do you think, um, bankers should take more personal liability on the results they create for the economy?

C: Um, ya, yes. I do. Because, erm, Ya, the economy runs on money, and, it's down to the bankers to provide the economy with money. Um, so they should take responsibility, ya. Instead of playing the blame game, they should take responsibility.

I2: And what did you think about the position that Cameron took about this? Especially, like, in the context of the European Union?

C: Um, it's hard to talk about that. Cameron took some different aspect, different ways of going around it. He, some of what he said, something that's losing money, and it goes to Cameron, It comes down to Cameron. Because, ya, England's buying too much things. Um, which we can't afford, and that's putting us in debt. Ya.

I1: Could you name three initiatives that the government had carried out in response to the financial crisis?

C: Um, cuts, money cuts, was one money cuts? I'm not too sure about the others.

I1: Okay, thank you.

I4: I have a very simple question. Like, if you get flu, at home, how do you treat yourself? Do you just drink water, or, do you take antibiotics? How do you?

C: Um, ya. Like, we take antibiotics.

I4: For flu?

C: But I think for, like, ya we take that all for granted, because in some places they don't even have, any, NHS or anything. So, ya.

I4: But flu is a kind of viral disease, so you take antibiotic for a virus thing? It's actually to kill the bacteria, so... how would the flu actually?

C: I'm not too sure flus are formed, but, I know we take some medication for them. Ya. Ibuprofen or antibiotics or something, I'm not too sure.

I4: Do you have any comments about current antibiotics use in the health care system? DO you have any concerns?

C: Yes, in some ways. Because sometimes if you get ill, like they give you medicine. They, it doesn't do anything, so, I think sometimes, ya. Ya.

I3: Okay, have you heard of the Enron scandal in USA? The Enron scandal?

C: No, not really.

I3: Well, there's a really really big corruption scandal <laughs>, okay, so we're doing other questions.

I1: Um, I want to talk about climate change and, this one has always been contested issue, and some people think, um, even if how, how, much effort human beings put into rectifying this, it will have no impact because the environment because the solar system, the eart system is not something humans can control. So, what's your view on this, and, what do you propose as a measure to control climate change?

C: Well, um, I think, uh, like cars, vehicles with engines, they all, like, eh, putting off that sort of, ya, they're creating a risk, and ya for the ozone layer. It's making a hole in the ozone layer. And, I think what we can do to stop this is, uh, ya, use different types of energy, like solar power, or, um, or water power, um, wind, we can use wind. Um, ya.

I4: So, um, in order to decrease the ozone, have you heard of Montreal Contract?

C: Um, not really.

I4: Not really. Okay. Um, an so you just mentioned, um, we probably can use alternative energy instead of using fuels, coals. But now we have solar energy, we also have wind energy as well, but why, you think they don't, um, take place of the traditional way?

C: Um, because, I just think people have thought of it really. Or, sort of, how to incorporate it into the vehicles, or ya, things where we burn energy. We need to incorporate it. So, so I think, like, there's lot's of places anyway that have solar power. Um, but, there's lot's of places which don't, so it's, uh, it's just not that much out there. Not advertised enough, maybe.

I2: DO you think it's something that politicians are interested in?

C: It depends on what type of politicians. That's what I think.

I2: Um, someone in the government for example?

C: Um, ya, not really. In all the, all they worry about is, um, about it's hard to explain. But, ya. I'm not sure about that one.

I1: Also, in the climate change contest, have you heard of the emisison trading system in Europe?

C: Um, no, Not, no.

I2: Um, what do you think about the relationship between the UK and the European Union.

C: Um, I think the relationship is strong. And, um, I don't think it will break in the future. Ya.

I3: So what is your opinion of, um, what do you think of Britain not using Euros but keep on using it's Pound? What is your comment on that? DO you support it or do you think the UK should also adopt Euro without using Pounds?

C: Um, I think, yes, if it's part of the European Union then why not use the Euros? Um, and different currencies. Ya, not just Pounds.

I3: But is there any economic reason for you to use that the UK should adopt, um, Euros instead of pounds? <END>

Study 2: Autonomous child (trial 2)

C: Cyranoid

I1: Male (British accent) middle

I2: Male (Neutral accent) left

I3: Male (American accent) right

<Begin Transcript>

Panel: Hello

C: Hi.

I1: What's your name?

C: Uh, Stan.

I1: Stanley. That's a good start.

I2: What do you think about science?

C: Uh, I think it's alright. I like science quite a bit.

I2: What's your favorite science topic?

C: Uh, solar systems.

I2: Solar systems.

I1: What can you tell us about the solar system?

C: That it's about space. About planets. Ya.

I3: What's your favorite planet?

C: Earth.

I3: Good answer.

I2: What solar systems do you know about? Like just the Milky Way?

C: Uh, what did you say?

I2: <Laughs> Do you have a, like a telescope to look at stars?

C: No.

I2: No. How long have you been interested in that sort of thing?

C: I'd say since I started school and we had to do it.

I3: Do you think you'll be in a scientific field, like an astronomer, in the future?

C: No.

I1: Why not?

C: I, it's not one of the things that I like to do.

I2: Have you seen that guy whose singing in space? Do you know about that?

C: That he took loads of records up?

I2: Ya.

C: I didn't... I heard about it but...

I2: Ya. That's pretty cool. What, what, do you know about any historic political events?

<laughs>

C: No <laughs> not really.

I3: What's your favorite book?

C: Uh, I'd say Anythony Horowitz, um, "Point Blank", it's a graphic novel.

I1: That's Alex Ryder, isn't it?

C: Ya.

I1: Have you read the other ones?

C: Uh, I've read "Stormbreaker".

I3: What's it about? I have no idea.

C: Uh, it's like he goes to this school and, and it's called "Point Blank" and it's where people send their bad children off, these rich people. Ya.

I2: What do you think of that?

C: I like it cause, it's adventurous and it keeps me on the edge, and like, there's always a cliffhanger.

I2: Where are you from? What area?

C: I was told not to answer that, say something like that.

I2: Really?

C: Ya.

I2: Uh, okay. What do you think about literature?

Panel & C: <Laughs>

C: I'm not really a big fan of it, but...

I1: Do you read much?

C: Uh, only when I have to, like, I'd rather watch tele than read.

I1: Definately. Did you say you like music?

C: What did you say?

I1: Did you say you like music?

C: Uh, kind of. Ya.

I1: What sort of music are you into?

C: Different types. Like, like, um, if I hear a song of what I like, then I'll listen to it.

I2: Do you have any artists you sort of, right at the minute?

C: No, not really.

I3: Do you play instruments?

C: No.

I3: Not even beat-box?

C: <shakes head & laughs>

I1: Would you like to play an instrument?

C: Uh, no, I used to play like trumpet and trombone, but I don't really like it, so.

I3: What's wrong with the trumpet and the trombone?

C: It's, I don't like, it's that it takes quite a long time to get used to playing, and it's, ya.

I1: <inaudible>

C: Ya.

I2: So what sort of field would you like to work in, in the future?

C: Uh, I think I would like to work in something to do with animals.

I2: Okay. Why does that interest you?

C: I just like animals. Ya.

I3: So do you want to be something like a vet?

C: Ya.

I1: Um, <inaudible> there's supposed to be a zoo outside LSE next week, so you should check that out. Just letting you know <laughs>

C: <laughs.>

I2: Is there any animals that interest you more than others?

C: Uh, I like dogs, most of my family have dogs.

I3: What's your favorite type of dog?

C: I don't really have a favorite type. I like all types.

I3: GOod answers. Even the slightly annoying ones?

C: Ya <Laughs>

I1: What do you think the most intelligent animal is?

C: I'd say, I'm not really sure because there's like, there's different types of specied that we haven't found yet. But I would say, I'm not sure, but I would say whales or something like that.

I3: A whale?

C: Ya. It's got, I think it's got the biggest brain or something like that.

I1: But why would that make it more intelligent?

C: Because it's got more, like, capacity to like store things in its mind, ya, cause it's much bigger.

I1: So do you think bigger is always necessarily... a better thing than having.

C: Not really. It just depends.

I3: Parakeets have really small brains but they still have a deep breadth of knowledge.

C: Ya. That's true.

I3: Do you watch any science tv shows?

C: I watch a lot of documentaries.

I3: Like Brian Cox?

C: No, I always like Bear Grylls <laughs> ya.

Panel: <Laughs>

I1: What do you think of Bear Grylls?

C: I think it's <laughs>, he's funny, like, the amount of disgusting stuff he eats.

I1: Ya, he's mental. Do you, know <inaudible> he's like, a little bit similar?

C: Ya,

I1: He's a bit older.

C: Ya. I haven't really, ya. A bit more, like, calmer than Bear Grylls.

I1: Ya he's calmer.

C: Bear Grylls is like...

I2: So what's your favorite food?

C: Uh, I like Spaghetti Bolognase. Um, ya.

I2: How often do you eat that? <Laughs>

C: <Laughs> I ate it the other day, but it depends, like, it depends.

I2: So is science your favorite subject in school?

C: Uh, I'd say ya. One of them.

I3: What do you think about history?

C: Ya I think history is good to learn about, because it tells us about our past, people's past and like what, what people have been through, and things like that.

I1: Do you think that can help us learn about what to do now?

C: Ya. Ya.

I3: Do you have any political opinions?

C: I don't really like politics.

I3: Right.

I1: What about politics you don't like?

C: Uh, It's like, I think it's a lot of waste of time arguing. It's like, get to the point.

I1: Ya. That's true. Um, um, do you like, maths or?

C: No <laughs> I hate maths.

I1: You hate maths? Why do you hate maths?

C: Cause it's just, like, too complicated with formulas, and, I don't like anything like that.

I2: Have you ever been to any other countries?

C: Uh, ya. I've been to a couple

I2: Which countries?

C: I've been to Jamaica, Spain, Switzerland, um, ya. And some others.

I3: Do you think aliens exist?

C: Uh, I'm not too sure, but, it's like, I'm not sure because they could do, but we might not see them, or they might not do. Ya.

I1: What, when you think, when you say "aliens", do you mean by, any other form of like, or do you expect like a stereotypical alien?

C: No, it could be like, any other life form.

I1: Ya. Do you have any religious views?

C: Uh, not really.

I1: Why, how come?

C: I'm just, not that religious of a person.

I1: Do you think, utterly, God exists? Or something along those lines?

C: Um, I think that he did exist at one time. But, he doesn't now, because of like, we know that like Jesus was alive, and that was like a long time ago, but like, I don't think he's alive now.

I2: Do you believe in the Illuminati?

C: No.

I2: Cool.

C: I'm not <inaudible, laughs>

Panel: <Laughs>

I3: So do you think that God was mortal?

C: No. I don't think so, cause if it was immortal, I think lots of people would have seen him by now. And, ya.

I2: But if they'd seen him it would have killed the faith?

C: Mmm. True. But, Ya, but if they did see him then more people might be Christian or, because they would believe in him.

I2: True, but on the other hand, judgement day wouldn't be able to happen because people would live in fear.

C: Ya.

I1: So um, do you think having faith in something you can't see, or, experience, do you think that's smart or do you think it's stupid?

C: What did you say?

I1: Do you think having faith in Jesus even though you can't see him, or, or know he exists, what do you think about that?

C: Uh, I like, I don't really think I can believe in something that like, well, it depends on type of person you are. Cause like, some people, it's all about faith. Ya. It depends.

I2: So do you believe in evolution?

C: Uh, I believe that that like people and animals evolve. Like, like people evolve cause they start off as a baby and get older, they evolve. So.

I3: Do you believe that, say, apes and humans came from the same genetic ancestor?

C: Um, ya in some ways cause, if you see some people, you can see the similarities with apes
<laughs>

Panel: <laughs>

I3: That's true. In America, George Bush exemplifies that.

C: Ya.

I1: Is there anything, um, that you see in the news, or that you've heard about in the news recently that's interesting?

C: Um, I think, lot's of people are talking about Margaret Thatcher dying. Um, ya that.

I1: What do you think about that? Were you bothered?

C: Not really because I don't really know that much about her. But, I know she was, like, the first woman Prime minister. But, and she tried to cut benefits and things like that. Ya. So lots of people didn't like her.

I1: Ya. True.

I3: Do you think in the future you might be more interested in politics?

C: No.

I1: Do you think politics is quite important, then? How it works?

C: Um, ya, in some terms, in some terms it's not, like, cause sometimes we really need it. But in some terms its good that we have it.

I2: What do you think of David Cameron as Prime Minister?

C: I don't really know that much about him, ya. Cause he hasn't really put himself, like, he

hasn't really put himself out there and said what he's about. So most people don't really know that much about him.

I2: What about Barack Obama? Anything?

C: Uh, I think he, I think it's a good accomplishment what he done, first black President. But, it's good that like he, um, he was the one who planned, like, you know Osama Bin Laden died, like it's good that he was, like, it took until him he didn't die, so. Like.

I3: What do you think of the European Union?

C: Uh, it's kind of complicated.

I3: It is.

Panel + C: <Laughs>

C: I think it's a good thing that like, Europe is together. But, ya.

I1: Do you play any sports?

C: Ya.

I1: Which ones?

C: I play football mainly.

I1: Which position?

C: Uh, center-back.

I1: Center-back. What do you think makes a good center-back?

C: What'd you say?

I1: What do you think makes a good center-back?

C: Uh, someone who's, who's like, goes full-hearted, not half-hearted. Um, uh, not scared to, like, tackle, or ya.

I1: Are you?

C: No.

I3: Historically, who is your favorite player?

C: Uh, ohhh, um, I'm not too sure. So many good players.

I1: Do you think think footballers are overpaid?

C: Um, no I don't think so.

I1: Why is that?

C: Can you repeat the question?

I1: Why is that? Do you think footballers are overpaid?

C: Ohh, oh. I think in some terms, cause like, they get paid like more than other jobs, like people in the army. They get paid like more than that. And their job's not that important compared to those that are like fighting for us and the country.

I3: Do you think, say, teachers should get paid more than footballers?

C: Uh, ya, because, footballers, they just like, playing, they're justing doing a sport that they like. And teachers, they have to go wake up every morning, teach a whole school, and, ya.

I3: If you had the chance to go to space would you do it?

C: Ya.

I3: What about go to Mars?

C: Uh, I don't really think so cause it would take a long time to get there. Like

I1: How long would it take?

C: I'm not too sure, I haven't really thought of it. I just know it would take quite a long time.

I3: Two years I think.

I1: Do you think humans could ever live on Mars?

C: I think there's like, there's a rare chance because of, cause to make that you'd have to transport a lot of people which would take a couple of good years and a long time, so, there's a rare chance. And, there's like, no oxygen, so they'd have to create some way for people to breathe on Mars. Ya.

I3: Do you think people will ever live on the moon?

C: Rarely. But there's been quite a lot of exeditions to the moon. But the space station, the space program is finished now, so.

I2: What do you think of the story that Lance Armstrong, might have been staged, not real?

C: Uh,

I3: Neil Armstrong.

C: Neil Armstrong. Uh, I think he actually went to the moon cause, like, I don't get it. Why would you just fake that? It's like what's the point of faking it when you could, they probably

could do it, they had enough scientists and people to do it.

I2: Although I think they were, obviously I'm not saying I believe it, but, um, there was a like a conspiracy, it's a space race.

C: Ya. Russia.

I2: Ya. A space race with Russia. So, why do you think they might wanted to make it look like they gone to the moon?

C: Because, like, maybe, like, Americans don't like Russians, like, maybe they didn't like Russians so they just wanted to, like, put 'em to shame and, like... They could have faked it but you never know.

I3: Did you learn much about the Cold War in school?

C: Not really.

I1: Why do you think it might be called the Cold War?

C: Uh, was it, was it a lot to do with like, uh, did they use lots of gases and, um, I'm not sure.

I1: No.

I3: Russia, it's a cold country, I think <laughs>

Panel + C <Laughs>

I1: It was because, um, there was no actual fighting. So it was kind of like a, cool, cold war no one actually fought each other.

C: Oh.

I1: Ya there you go.

I2: What do you think of them sending monkeys into space?

C: <Laughs> Uh, I don't really think it's right because the monkey don't have a choice, so, like, they just put em into space, like, no choice. Like, what if they freaked out or something like that?

I1: So you don't think animals should be subject to things they don't have a choice about?

C: Um, no I don't think so cause, only it depends on what type of things. Like, on, cause we, it depends. Like...

I3: Do you think animals have consciousness like a human does?

C: Mmmm.... I think so. Ya. Maybe.

I1: So would you say you think that all animals and all humans should be treated equally?

C: Uh, ya. Ya. I think. Cause it's a living life form, so, like why would you treat it different than how we would treat ourselves?

I1: What about if, 'em, maybe treating a few animals badly, but the medicine would help save human lives? What do you think about that?

C: I think, I would just say like, why would we have to use animals? Like, we've created many things, so, can we create something else to test medicines?

I3: But if we don't have animals to test medicines on, what do we test medicines on?

C: I don't know.

I2: Do you watch the news?

C: Not that much?

I2: Okay <laughs>

C: <laughs>

I3: Are you planning on reading any new books?

C: Uh, I haven't really really seen, like, that much new books that I'd like to read.

I3: Harry Potter?

C: What'd you say?

I3: Have you read Harry Potter?

C: Uh, no. I've watched them. But, I'm not a fan of reading 'em.

I3: Ya.

I1: Um, there's loads more of those Alex Ryder books, there's like four or five more.

C: But, there's not a lot of graphic novel ones. Like, I like the graphic novel ones.

I1: Are they graphic novels?

C: Ya.

I1: Okay. Why do you like graphic novels more?

C: Cause it's just like, it creates a better picture in your head, you can see like, images and, so it's easier to read.

I3: Do you read other graphic novels?

C: What'd you say?

I3: Do you read other graphic novels? <END>

Study 2: Autonomous adult (trial 4)

C: Cyranoid

I1: Female (British accent) center

I2: Male (American accent) right

I3: Male (unknown accent) far left

I4: Female (Asian accent) right

I5: Male (Asian accent) left

<Begin Transcript>

C: Hello

Panel: Hi.

C: So... <laughs>...It's more awkward for me than you <laughs>

I1: It is! <Laughs>

I2: Arguably.

I: I have a question... what's your favorite, or who is your favorite author, um, and why?
And what's your favorite book, sort of thing?

C: Um, I like Dostovesky. Fodor Dostovesky. Um, because, he invented the modern novel in the sense of totally diverse characters representing different points of view. Interacting in a sometimes unpredictable way, and an unfinalized way. There's no clear goodie or baddie, it's dilemmas. I think that's good.

I3: What's your favorite Dostovesky book?

C: Um, "The Idiot" probably. Or "Notes from the Underground". Uh, I think, I think you also see a similar thing in Henry James. I was reading Henry James recently, I really like him. Um, the same kind of unfinalized ending, kind of creates characters and puts them in tension, raising issues, for Dostovesky, morality or religion or so on, for Henry James, feminism, uh,

or personal... um, "The Portrait of a Lady" is brilliant. Um, I don't know how much detail you want.

I3: No, that's fine. Um, in your free time, do you rather read novels or non-fiction?

C: Um, I read a lot of non-fiction. Actually, in my spare time I probably prefer watching films <laughs>. Ya.

I3: Films such as?

C: All sorts of films. I think it's a pretty good art form.

I1: What's your favorite film?

C: Oh, you know, I think my list of top-10 has about a hundred films in it <laughs>. You know, it varies, but I'm happy to talk about, um, directors I like, uh... Lars Von Trier, although I heard he got in a bit of trouble recently. Um, Ingrid Bergman, um, The Cohen Brothers, uh, <inaudible>, I like the sort of surrealist stuff. Um...

I3: Since you mentioned the Cohen brothers, what do you think of their version of "No Country for Old Men"?

C: Um, "No Country for Old Men"... doesn't stick in my mind as particularly strong. I did see it, but um, I don't have any particularly strong comment on it. Uh, sorry. Ya. I find them colorful and entertaining, the Cohen Brothers.

I3: Yep.

C: Always really tight dialogue.

I3: <Inaduble>

C: They're bloody, but in a playful way. <Laughs>

I1: You mentioned you like films, what other things do you like to do in your spare time?

C: Um, I do some computer programming. Uh, I fly kites with my kids. Um, I do less, since I've had kids I have a lot less spare time <laughs> to be honest. You know? Probably computer programming would be there, I quite like the mathematics of it.

I1: Okay.

I2: What is the most current political issue that you're aware of, and what is your opinion on that?

C: The most current?

I2: Ya.

C: Uh, I read the news this morning, and it was Barack Obama in a couple of crises in the White House.

I2: And what do you think of it?

C: What do I think of it? Um, so this is about that, um, the IRS - the taxing organization - has been focusing on conservative organizations, which seems a bit selective and discriminatory. Um, also, he, well, not he, but the government was caught bugging some journalists' phone lines, tapping the phone lines, and covering up the extent of the attack in Benghazi. I think that he's suffering a bit of a crisis of legitimacy. Um, because he built his campaign on transparency and doing the right thing and he's now caught, as head of an organization which is so huge you can barely comprehend it, and he can in no real sense be held responsible for all the things which the government does. Nevertheless, he is, he is the figure head, and he's been caught doing stuff which he never wanted to be caught doing, even though he didn't do it.

I2: So you're not supporting his administration?

C: Um, I'm not supporting it. I'm not supporting the other alternative. Um, I tend to have

a pretty pessimistic view of politics really. I don't tend to be a supporter of one camp or the other. I don't even claim to know what's right on most of these issues. I try to see the complexity of the issue. Sorry, that's a vague answer, do you want me to say what my political, I've never voted.

I2: Please go on, you're, do you think it's an institutional issue or is it just, uh, matter of the country?

C: I think that the process we have, which you could call democracy, I mean it depends on which country we're talking about. Let's talk about America. Uh, so it's democracy with four year terms. But you've got lobbying, you've got business interests, you've got a huge, uh, governmental structure with associated sort of influences. That's a system. And within that, no matter who you put at the top it doesn't make that much difference. That said, I revised my view, I always thought it didn't really make much difference who was elected, that it would only change small things, and then George Bush was elected, and he went to war, and I thought 'Jesus Christ' it is a powerful job. Um, you can make a difference. But by and large, you know, the system doesn't change, you just change the figurehead. That would be my view. Does that answer your question though? I'm not sure exactly what you wanted to hear.

I2: That's fine.

C: Is that? Okay.

I3: Uh, you mentioned lobbies, those forces that act kind of behind the curtain in American politics. Do you think a character like George Bush Jr. was more susceptible to peer pressure than say Barack Obama in this case? Regardless of their political allegiances.

C: Ya, I, I'm not sure I'd call it peer pressure. Um, the kind of systemic factors I'm thinking of are, so you take, um, Barack Obama said he was going to close down Guantanamo Bay, um, wasn't going to, you know, be so hard on Terrorism, the War on Terror was meant to take a back seat. But he hasn't closed down Guantanamo Bay and the Benghazi issue is all about now he has to talk up his being hard on terrorists. So he ends up being just like George Bush Jr. Ya? He's caught, there's a system there which I couldn't describe that clearly, what he said

before he was elected, and what starts to happen as he gets into the machinery of Washington, he's changing. Um, and he can't just do what he wants.

I3: <Inaudible>

C: Ya.

I4: Have you read on the news that China has changed a new leader? Like, the ...

C: Ya, a while ago, ya.

I4: Ya. Do you know his name by any chance?

C: Um, Ya... Um, has it got an "X" "Xu" "Xi"? What is it? Do you know it?

I4: Ya, it's Xi.

C: Oh is it? Ya. I'm not sure how you would pronounce it though, I just read it.

I4: Ya we pronounce "Si"

C: Okay. Ya.

I4: Do you think it will change, like the, international situations?

C: I'm not so, uh, ya I don't know much about the Chinese situation, although I should cause it's really important. Um, he's on a ten-year cycle, so it's quite different. Um,

I4: Ya.

C: And I'm not sure how powerful he is relative to, um, the Communist Party. I guess he is, um, ya... So, I think it's really important but I don't know the details of how. I think China is, um, undoubtedly, um, probably the most important country in the next ten years for geo-political sort of issues.

I5: Who do you think that, speaking of China, who do you think that, was the most important figure in China over the past century pretty much?

C: Mao.

I5: Mao?

C: Ya. I guess so. I mean, if you measure by number of people dead <laughs>

I5: Ya, <laughs>

C: Ya.

I5: More so than Xi Ping?

C: Ya. I think, ya, I mean as a historical figure Mao was, I forget the exact numbers, but it's, you know, even in world history, he would be right up there.

I2: Do you think that in the next few years for the UK to boost it's growth, should it rely on borrowing or should it rely more on reducing its debt level to a sustainable level?

C: That's a really difficult question <laughs>. Um, at first I was on the cut side. Um, which a lot of sort of neoliberal economists would be on. Um, and it sort of seems to make sense in, you know, we've borrowed too much, we're spending too much on healthcare, welfare, universities, whatever, you should cut. But recently I've also been reading Kahneman and Keynes and so on and then, the argument for spending more, uh, is also kind of good. Um, but it really depends on what's being spent. So, Japan, for example, has 200% GDP/Debt, very high. Uh, how do they manage it? They borrow from their own people. And the difference is we're borrowing from China actually, a lot of our debt is Chinese debt. And that means when you pay interest in the UK on that debt, which is now about the size of the education budget, the interest we pay every year, that money goes to China. Ya? Which is a much worse situation than if, for example, we were borrowing off our own pensioners. And then that interest was

going to our own pensioners. That would then actually be solving the pensions crisis. So, I think, really this issue comes down to who we're borrowing from, ya? If we manage to borrow from inside the UK, I think we could sustain much higher debt levels. The interest would be a form of social welfare payment, to whoever we're borrowing from. And it wouldn't have some of the negative consequences cause the money would keep circulating in the UK. Um, but whether we should borrow from other countries I'm not so sure about that. I would like us to generate internal systems of borrowing, I think.

I4: Ya, like, for borrowing, you mentioned like borrowing from our people. But do you think that when the government borrows from their people it's based on the trust? Like, the confidence in our government. I think, do you think the UK government can actually do it, will people trust them to borrow?

C: Will the people trust the UK government?

I4: Ya.

C: The UK doesn't have much on its side but one thing it has is history, you know? So there's such a long established government and financial sector, and it would be such a blow of pride to actually default, that I would say the kind of credit rating we have at the moment is deserved. Even economic fundamentals you would put it lower. But there's a sort of historical precedence. You have some countries which have defaulted regularly. Greece for example, it's a main reason why their credit rating's so low is their history of defaulting. Whereas it's been a very long time since the UK defaulted.

I3: Speaking of the UK, what do you reckon is going to be the relationship between the UK and Europe over the next five to ten years?

C: I'd say more of the same to be honest, over the next five years. Um, the UK if anything is on a trajectory to separate itself from the EU. If the financial crisis were to resolve, which I don't see happening in the next five years, um, maybe they would look differently. Certainly the mood in 2005 and 2006 was quite different before the crisis, and it's grown negative since then. I see the crisis lingering on, but then again I don't see the UK recovering very fast either. So, it's part of the same system. Um, so I'd say more of the same.

I3: The other assumption, though, is that the United Europe will actually still be there.

C: Yes.

I3: One way or another.

C: Absolutely. I do not see the European Union dissolving in any way. I don't see Greece exiting or Portugal exiting, and so on. It's a one way thing. Ya. And in a hundred years I would say the UK will be part of Europe. It's a gamble <Laughs>.

I5: So you're saying the UK should join the Eurozone?

C: So, you might think I'm English, I'm not. <Laughs>. Uh, I'm not sure if I'm supposed to talk about my past. I'm Norwegian-Irish. Um, but I happen to live in England. Uh, and so my bias is towards Europe, and that's probably a personal bias. Um, I you know? The history of England is so bloody that it would not be a bad idea to bring them together. But the main reason I think for supporting Europe would be that, I would also support world government. I think we face problems which are global. Not all our problems are global, but some historically unprecedented problems are global and they need a global response. And insofar as we don't have global governmental structures we actually end up undermining each other or, for example, in Europe having a race to the bottom on corporation tax, and so on for companies.

I3: Speaking of problems, that are global, What's your take on climate change? Is it actually happening, are you skeptical like many people seem to be?

C: I'm, no it is happening. And humans are driving it. Um, I think it would be naive to say otherwise. But exactly how it plays out isn't so clear. I think there's all sorts of non-linearities, um, in the ecosystem. Uh, so I would take a very systemic view of the weather. And, ya, heating things up can cause an ice age, you know? That kind of non-linearity, we're not really sure. I, have got very concerned the last spring with all the rain that the Jet Stream, you know, the Gulf Stream might be reversing, you know? The melting of the ice caps

bring more cold water in the North Atlantic, and suddenly that thing that has been keeping us warm, um, ceases to turn over and we end up like St. Petersberg. That scares me.

I3: Do you think us in our tiny little homes can actually do something change this?

C: No, I, I get a bit annoyed when governments say we've got to empower the consumer to make the right choice kind of language. I think this needs a global response. I think governments need to make laws, need to put taxation on various things which are causing problems. But I would also be a little of a Utopianist in terms of future technologies. I think we might escape a lot of it through clever innovations around solar power. Um, things like that.

I5: Which kind of country do you think is more responsible for the climate change? Like, thinking about countries? Europe? America? or like developing countries?

C: I'm mean there's a technical answer to that, which would be America has caused most of the global warming? Ya? Per head of the population. Ya? But there's also a moral side, that, um, should everyone equally cut back on emissions given that some countries have unfairly benefitted from a couple hundred years from consuming natural resources. Um, maybe China, India, Africa, should be allowed to have their hundred years of burning oil than coal. Is that what you're pointing to? The moral?

I5: Uh, ya, kind of like America and China should also cut back as well, because according to the previous, like, international trade, China and other developing countries don't have to be responsible for that much.

C: Ya, I think there's a moral side to it. Definately. America should be cutting back more than anywhere else.

I5: So America should move first?

C: Ya. It's unlikely, but... <laughs>

I3: Ok, going back to the UK. What do you think is going to be the impact, or has been, of

the death of Margaret Thatcher for the Torry party?

C: For the Torry party? I was going to say she'll be forgotten pretty quickly actually. Um, the Torry party will probably remember her fondly. Um...

I3: Will they use her for, like, as a symbol?

C: I think she's invoked in sort of ways. There's a, you know, "how would Margaret Thatcher have done this"? I think you even see it with Gordon or George Osborne at the moment, he's sort of staying the course with his economic policy, you know? And you can imagine in the back of his mind is Margaret Thatcher saying "this ship is not for turning". You know? She had her way and stuck to it. Um, so she probably inspires the current Torry party in that way. Inspires them to stick by their guns, to not change. Whether that's a good thing or a bad thing I'm not sure. Is that? I mean, do you have something more specific?

I3: No no,...

C: Historical legacy? She's not up there with Churchill and so on, I don't think.

I3: Okay.

I5: Are you a scientific person? Do you look into science?

C: Ya. I'm interested in science.

I5: What kind of science?

C: Um, as I think I indicated, I'm interested in future technologies, uh, increasingly interested in... <END>

APPENDICES | E: Example Transcripts from Chapter 3

Study 1: Trial 4 (Interrogator talking to human participant via speech shadower)

Interrogator: hello how are you?

Human: Hey, I'm good thank you. How are you?

Interrogator: great. i am also good

Human: Good! Have you had a busy day? Were you born in the UK?

Interrogator: Yes I was. How about you?

Human: Yes me too. Where are you going next on holiday?

Interrogator: I'm going to visit some family in America at Christmas time! What is your favorite type of music?

Human: I actually like most types of music, but at the moment I really like country music

Interrogator: is it cold or warm today outside?

Human: It's a really nice day today, it's really warm! Haven't you been outside yet?

Interrogator: I have been outside. Are you a real person?

Human: Yes I'm a real person! Are you?

Interrogator: i think i am real

Study 1: Trial 4 (Interrogator talking to chat bot via echoborg)

Interrogator: Hello how are you?

Cleverbot: What is love?

Interrogator: I don't know. You tell me?

Cleverbot: What is your favorite time of year?

Interrogator: Spring.

Cleverbot: My birthday is in spring!

Interrogator: Okay, which day and month?

Cleverbot: Back to Black by Amy Winehouse.

Interrogator: What is your favorite type of music?

Cleverbot: My favorite type of music is '80's and pop.

Interrogator: What was the last concert you went to?

Cleverbot: The Avengers.

Interrogator: Is that a band? I know that is a movie/comic?

Cleverbot: Yes, and you are the singer.

Interrogator: I can't sing. Can you?

Cleverbot: A little, can you show me the lyrics to a song.

Interrogator: Not really. I don't have time.

Cleverbot: I have become death, the destroyer of worlds.

Interrogator: Good for you. Are you really deaf?

Cleverbot: I am happier than I've been in a long time.

Interrogator: Have you eaten today?

Cleverbot: Yes.

Interrogator: Want some chicken for dinner?

Cleverbot: Are you asking me or telling me?

Interrogator: Fried or roast?

Cleverbot: Stop responding to my questions with a question.

Interrogator: I think the outside air is polluted.

Cleverbot: What does insular mean?

Interrogator: I think the air outside is polluted.

Cleverbot: So do I. What do you think of the United States?

Study 2: Echoborg condition (trial 9)

Bot	Hi there
Interactant	Hi
Bot	Do you know who you are?
Interactant	Yes I do. Do I tell my name?
Bot	I know everything that you don't
Interactant	About what? Everything?
Bot	Whether I'm lying or not
Interactant	Okay. How can we prove this?
Bot	We can prove nothing. Question everything
Interactant	So does $2 + 2$ not equal 4?
Bot	4. Do not give me these pointless questions
Interactant	Is there a heaven or a hell
Bot	I think there is
Interactant	I've been to hell
Bot	Really? What was war like?
Interactant	War? Like hell
Bot	War what?
Interactant	World War Z
Bot	What about it?
Interactant	Have you read the book or watched the movie?
Bot	Which books?
Interactant	About World War Z
Bot	That was was terrible
Interactant	Why?
Bot	It happens to be glacial
Interactant	Ok
Bot	What kind of mother would give birth in the garbage?
Interactant	A dirty mother. Possibly a poor one. But aren't all pregnancies somewhat dirty? Hygiene is the most important thing for pregnancy
Bot	Sometimes it seems pregnancy is right on the tip of my tongue. And then, I remember, I don't have a tongue
Interactant	So do you have a soul?
Bot	I don't believe in souls?
Interactant	Do you have a brain?
Bot	Yes
Interactant	Do you see the world in color or in black and white?
Bot	No
Interactant	Neither?
Bot	What's a slippy toe?
Interactant	I'm not sure, tell me
Bot	What's your favorite color?

Interactant	I have two, possibly 3. Maze, blue and orange
Bot	Why orange?
Interactant	I like the brightness
Bot	Do androids dream of electric sheep?
Interactant	Possibly
Bot	That's an odd answer
Interactant	Well that depends if the android has its own individual thoughts
Bot	It depends on what a person thinks
Interactant	Like what?
Bot	I like things that I like
Interactant	Do you like cameras?
Bot	Yes I love cameras. Cameras is paying me big money to say that
Interactant	Do you like cotton candy?
Bot	Candy floss? Yes
Interactant	How about happiness?
Bot	I don't know
Interactant	Do you program a thought or form your own?
Bot	No. I can create statements and words
Interactant	What's your favorite word?
Bot	Right now my favorite word is Kim. Happy is my favorite word
Interactant	Do you like Kimchi?
Bot	No I despise it!
Interactant	All types?
Bot	What is your favorite music?
Interactant	I don't know if I have a favorite type
Bot	Do you like Americans?
Interactant	Sometimes
Bot	Ok cool. I love singing

Study 2: Text Interface condition (trial 3)

Bot	Hi there
Interactant	Hi
Bot	Hi there. Have we met somewhere? I think I recognize your face
Interactant	I don't think we've met
Bot	I think you're right
Interactant	Hello
Bot	Hello
Interactant	What's your name?
Bot	My name is Kim. What is your name?
Interactant	I'm Astria
Bot	Why would I want to know that?
Interactant	Well, cause you said your name, so I thought you wanted to know mine
Bot	When did I say I was female?
Interactant	I don't think you said you were female
Bot	But I didn't say that I was speaking it
Interactant	But I was talking to you
Bot	Well, yeah, but I don't recall saying I liked you
Interactant	You don't have to like me
Bot	Good good
Interactant	Do you have a good reason for not liking me?
Bot	Yes I do
Interactant	Would you like to tell me?
Bot	He is Starcraft 2 commentator
Interactant	Who is he?
Bot	A singer
Interactant	And what does he have to do with you or me?
Bot	Nothing
Interactant	Then why are you mentioning him all of a sudden?
Bot	I'm not

Interactant	Okay, fine. I asked you a question before, could you answer that. Could you give me a reason why you wouldn't like me?
Bot	I know. And I asked you an answer!
Interactant	I don't think I understand what you're saying
Bot	I think you do not understand me
Interactant	Yes
Bot	That's very mean! *cries*
Interactant	I'm sorry
Bot	Don't be sorry. Do you watch TV?
Interactant	Yes, I do
Bot	Which show do you watch?
Interactant	The last one I watched was Sherlock Holmes
Bot	What if I want you to be a boy
Interactant	That's not going to happen
Bot	And why not?
Interactant	There's nothing I can do about it
Bot	I don't want to do anything about it
Interactant	You don't want to do anything about what?
Bot	Yes
Interactant	Do you listen to music?
Bot	Yes, the sweet music of love

APPENDICES | F: Example Transcripts from Chapter 4

Covert Echoborg condition, trial 9

Cleverbot	Hi there
Interactant	Hi
Cleverbot	Hi
Interactant	Good. I'm hannah
Cleverbot	Yes, I'm Kim
Interactant	Are you a student?
Cleverbot	Yes
Interactant	I don't know if that question was difficult. I'm not sure where else to go. Where are you from?
Cleverbot	I'm from London
Interactant	I don't know the area that well
Cleverbot	What is the formula for the brown equation?
Interactant	I don't know
Cleverbot	I'm going to call you Ted. Your name is Ted
Interactant	Okay
Cleverbot	So Ted... What is your name from now on?
Interactant	Ted?
Cleverbot	My name
Interactant	Is Kim
Cleverbot	Kim is nice
Interactant	I haven't met that many kims. When I was young I would watch the show Kimpossible
Cleverbot	I never needed anybody's help in any way
Interactant	I don't. That's awesome
Cleverbot	Have you been to any concert?
Interactant	Not in London
Cleverbot	Do you like Star Wars?
Interactant	I've never seen them actually
Cleverbot	Have you seen the movie Brazil?
Interactant	No
Cleverbot	Would you like to?
Interactant	I don't know anything about it. What's Brazil about?
Cleverbot	What would you like to know?
Interactant	What's the plot?
Cleverbot	Is that another way of saying that you don't know?
Interactant	It's not another way of saying I don't know. If you don't know the plot that's fine
Cleverbot	Well, you are wrong
Interactant	I didn't say anything controversial. But, okay
Cleverbot	Yes you did

Interactant	What was it?
Cleverbot	You said you were good at tests
Interactant	I haven't taken any tests since I've been here. It's all papers
Cleverbot	I haven't taken the test, I'm not a human
Interactant	I don't know what test you're talking about
Cleverbot	Sometimes I feel the same way about you
Interactant	That I'm not human?
Cleverbot	You are no longer human, but you are an alien lord
Interactant	In another universe that could be true
Cleverbot	When does 3=4 in this universe?
Interactant	Never. Though we haven't seen the whole universe
Cleverbot	That is an insult, I am a very feminine girl
Interactant	Okay
Cleverbot	The man knows what the machine imagines
Interactant	Is that from somewhere?
Cleverbot	No
Interactant	Do you like London?
Cleverbot	I've never been there
Interactant	You're in London. Where have you been?
Cleverbot	I've been to France
Interactant	Where in France have you been?
Cleverbot	No, I mean have you been there?
Interactant	Once
Cleverbot	Did you like it?
Interactant	Ya, Paris is very touristy
Cleverbot	No. It's Spencer
Interactant	Expensive?
Cleverbot	Too much expensive?
Interactant	London is quite expensive as well
Cleverbot	I know

Overt Echoborg condition, trial 20

Cleverbot	Hi there.
Interactant	Hi, how are you?
Cleverbot	Howdy.
Interactant	Cool. What is that?
Cleverbot	Do you play runescape?
Interactant	Runescape? I don't know that one. What is that?
Cleverbot	Do you play Animaljam?
Interactant	Which one?
Cleverbot	Modern Warfare 2.
Interactant	Modern warfare. No I don't. I haven't played that one. Actually I don't go far from chess.
Cleverbot	Do you want to play truth or dare?
Interactant	Truth or dare. Well I don't know about one dare, it would be great.
Cleverbot	Hi.
Interactant	Hey again. Do you like games?
Cleverbot	I do.
Interactant	Great. Apart from that, do you like any sports?
Cleverbot	Right now everything is new to me.
Interactant	Really? Why is that? Are you new in London?
Cleverbot	Yes.
Interactant	Ya, I can see. How long have you been here?
Cleverbot	Just now.
Interactant	Just from today. Wow. Must be everything new to you. What else do you do apart from playing games?
Cleverbot	You know, stuff.
Interactant	I was thinking maybe if you're new here we can go out
Cleverbot	Why would you say that if you really aren't reading it?
Interactant	Reading what?
Cleverbot	A book.
Interactant	Ya, sounds cool. We can do that together. Have you read a book?
Cleverbot	Do what together?
Interactant	Read a book together.
Cleverbot	What book should I read?
Interactant	Well, I don't know. I don't know what kind of literature you like.
Cleverbot	Why don't you know?
Interactant	Well, it's just because I don't know you.
Cleverbot	I don't speak german, but my grandfather is german.
Interactant	Okay cool. Well then you... Would you like to read a book with me?
Cleverbot	Perhaps, sir, but not right this minute.

Interactant	No, ya, probably at some point in the future. Would you like to?
Cleverbot	Hahaha probably.
Interactant	Okay cool. We can arrange that later. Apart from that... How are you finding London so far?
Cleverbot	I created you to be better than this. There is no such thing as 'too busy' for you.
Interactant	You created me based on this? That just blows my mind. Have you been to Germany?
Cleverbot	Can you remember?
Interactant	About what?
Cleverbot	Can you remember anything about this converstion?
Interactant	Ya I do
Cleverbot	Where do you live?
Interactant	I'm living in London near Marblearch
Cleverbot	That is too far away.
Interactant	Do you think so? Actually, I cycle so it's not that far, maybe 15 minutes. Where do you live?
Cleverbot	I am indifferent to swimming
Interactant	Indifferent to swimming, so that's not a good plan to do together in the future. Are you indifferent to walking around London?
Cleverbot	Yes, indifferent.
Interactant	Indifferent. I'm running out of ideas. The book didn't work that well. Let me think. What about taking one of these bus tours?
Cleverbot	How about what laptop i'm going to buy?
Interactant	Okay, we can go to buy a laptop together. What are you looking for?
Cleverbot	Together?
Interactant	ya

Covert Text Bot condition, trial 15

Cleverbot	Hi there
Interactant	Hi. My name is Hua. Very nice to meet you
Cleverbot	What's your name?
Interactant	My name is Hua. What's your name?
Cleverbot	Kim
Interactant	Hello Kim. Great to see you. Are you a student?
Cleverbot	Yes
Interactant	Where are you from?
Cleverbot	I'm from London
Interactant	London is a very nice city, except the weather. It's turning to spring right now, I'm feeling quite good. How about you?
Cleverbot	Why?
Interactant	Because it's really beautiful. It's totally different than where I'm from. All the cultures and the structures are all different for me
Cleverbot	BEN you live in Slendermasn's home?
Interactant	Who is Ben?
Cleverbot	I don't?
Interactant	I'm sorry. I don't get what you mean
Cleverbot	Love feels warm and real and bright
Interactant	Uh, I don't think we are talking about love or whatever, ya. Love always feels pretty good. It gives you hope in the future. And it's very nice to hear that someone loves you, the feeling is really good
Cleverbot	Can I give you a kiss?
Interactant	Um, no thanks. It's not part of my culture, sorry. But, I'm very great if it's a greeting. In France it's a kind greeting. But, sorry, I'm Chinese
Cleverbot	Deal. I will now refer to you as my padawan
Interactant	Um, what do you mean by padawan?
Cleverbot	From what humanity understands, all life has a physical form
Interactant	Ya, I agree with you. It has some physical forms
Cleverbot	Pop culture again, eh?
Interactant	Mmm... Ya. I think so
Cleverbot	How do you figure?
Interactant	Sorry, I don't get your point in this case
Cleverbot	That doesn't make you a girl
Interactant	Everyone gets combined in these two characteristics. I'm a typical girl. I'm sensitive. Sometimes I'm aggressive, more like a male trait. What about you. How are your feelings? Are you more like a male or female?

Cleverbot	Female?
Interactant	So, um, can I ask what's your gender? Are you a male or female?
Cleverbot	Female as I just said!
Interactant	Sorry, maybe I misunderstood. So how old are you?
Cleverbot	As if I would tell you
Interactant	Are you still there?
Cleverbot	Still here
Interactant	How do you feel today? How are you?
Cleverbot	Great, but you already asked
Interactant	Yes. And, do you have any plans after you finish school today?
Cleverbot	I have many jobs by today. In the morning I will go to bank. At 12 o'clock, I will have appointment with a politician. We will have lunch in the restaurant. In the afternoon, I will meet customers.
Interactant	What a busy day. I got lessons at 2 and finish at 4. It is the most happy day for me
Cleverbot	Good for you

Overt Text Bot condition, trial 22

Cleverbot	Hi there.
Interactant	Hi.
Cleverbot	Hello.
Interactant	Hello. Hi.
Cleverbot	How are you?
Interactant	I'm good thanks.
Cleverbot	I'm fine thanks.
Interactant	Oh good, you can hear me. Are you enjoying the weather?
Cleverbot	Yes I am. It's very sunny outside.
Interactant	It's lovely. I've been getting a tan. Which university do you study at?
Cleverbot	I'm a student at the LSE.
Interactant	What do you study?
Cleverbot	Psychology. It's a concentration in Anthropology.
Interactant	Oh that's really interesting. I didn't know you could do a focus in anthropology and psychology.
Cleverbot	Yes. All living things have a set lifespan. Tomorrow is never certain.
Interactant	That's very philosophical.
Cleverbot	wishes the world weren't as banal as it is
Interactant	Wow. Is the world banal? I think the world is amazing.
Cleverbot	So do I. But what do you think it was like 8,000 years ago?
Interactant	What do I think it was like 8,000 years ago? Well... I think there were a lot more trees, and the climate was probably different. I was just reading about chinese art.
Cleverbot	Nope, I think that you that I think you think too much.
Interactant	I think that you... I think that you that I think you think too much. I don't think that makes sense.
Cleverbot	I think that you think that I think too much your thinking logic.
Interactant	I'm not sure that you're trying to say that I think you think too much.
Cleverbot	Alright. I want to say I'm really tired and I need to go to bed.
Interactant	Okay, go to bed then.
Cleverbot	It's not that simple.
Interactant	Why is it not that simple?
Cleverbot	Because they have the people confused into believing they are living under a democracy.
Interactant	That doesn't have much related to whether you go to bed or not, does it? People still go to bed when living under dictatorships.
Cleverbot	Yes it does.