Special Issue: Decoding Artificial Sociality

new media & society
2025, Vol. 27(10) 5592–5621

DOI: 10.1177/14614448251338277
journals.sagepub.com/home/nms

The conversational action test: Detecting the artificial sociality of artificial intelligence

Saul Albert
Loughborough University, UK

William Housley
Cardiff University, UK

Rein Ove Sikveland
Norwegian University of Science and Technology (NTNU), Norway

Elizabeth Stokoe
The London School of Economics and Political Science, UK

Abstract

Drawing on the "Voigt-Kampff Empathy Test"—a science fiction version of Turing's famous thought experiment—we propose the Conversational Action Test (CAT): a new approach to evaluating conversational artificial intelligence (AI) voice agents. We compare social actions in a range of telephone service encounters where one party is an artificial conversational agent to a range of similar human-human calls. The CAT demonstrates a novel paradigm that addresses long-standing theoretical and methodological problems for ostensible "tests" of conversational AI by (a) revealing the conceptual confusion of attempting to "detect" an AI in routine service interactions and (b) focusing, instead, on the situated interactional practices through which an AI "passes" for human. We discuss the implications of the CAT for the design and evaluation of conversational AI, and for the notion of "humanness" as a goal or benchmark for such

Corresponding author:
Saul Albert, School of Social Sciences, Loughborough University, Brockington Building, Loughborough LE11 3TU, UK.
Email: s.b.albert@lboro.ac.uk

systems. Data include publicly available human/AI service calls and comparable human-human calls in British and American English.

Keywords
Conversation analysis, conversation design, conversational AI, conversational user interfaces, voice interfaces

Introduction

In Ridley Scott's (1982) film *Blade Runner*, the "Voight-Kampff Empathy Test" distinguishes androids from humans by monitoring the subject's biometric responses while the examiner describes a series of grotesque scenes. This interpretation of Alan Turing's (1950) thought experiment, fictionalised by Phillip K. Dick, imagines a future of ubiquitous "strong deception" in human-machine communication (Natale, 2023), in which it has become otherwise impossible to tell them apart. By 2019, the year in which the film's events are set, Google's conversational artificial intelligence (AI) agent *Duplex* (Leviathan and Matias, 2018) was able to mimic human callers well enough to make booking calls to real restaurants and salons, with the artificial agent apparently passing as human "in the wild" at its product launch demonstration[1]. *Duplex* has since been withdrawn amid questions about the ethics of its mimicry (O'Leary, 2019), its efficacy (Bonifacic, 2022), and, ironically, about the authenticity of its demonstration calls (Natale, 2021). Once *Duplex* was publicly deployed, with its automated agents beginning encounters with the preface: "Hi I'm Google's automated booking service" (Dwoskin, 2019), businesses apparently started ignoring *Duplex's* "spam calls" (Garun, 2019). This suggests that the functionality of these systems may hinge on the ability to pass as human. Though *Duplex* was discontinued, AI call centers now offer similar services[2]. The "deceptive AI ecosystem" (Zhan et al., 2023) that these systems now inhabit, enhanced by Large Language Models (LLMs), further enables AI agents to navigate a range of conversational situations. Given the challenges of detecting AI-generated text (Else, 2023; Liang et al., 2023) and much-vaunted claims that LLM technologies now "pass the Turing Test" (Adams, 2024), there are increasingly urgent calls for telephonic equivalents of the "Voight-Kampff" test (e.g. Shen et al., 2024).

In this article, however, we start by reconsidering what it means, in practical, interactional terms, for an AI to "pass" as human in the context of a routine service call. Natale (2023: 92–123) suggests that the "Eliza effect," named after Weizenbaum's 1960s ELIZA psychotherapist bot, not only biases us to ascribe agency to even the simplest bots, but also constructs a mediagenic narrative about the boundaries between humans and machines. Should we be developing tests for Voight-Kampff-like behavioral "tells" to disambiguate humans from AI? Or does the very concept of a test for humanness uphold a flawed narrative about human authenticity and sociality that, as in *Blade Runner*, dehumanizes both tester and subject? Here we rethink the notion of such a test in relation to sixty years of research in conversation analysis (CA). We contribute to an emerging approach to "conversational AI" that looks beyond common interpretations of the Turing Test as either a deceptive "imitation game" or as an operationalization of machine "intelligence" (French, 2000) by analyzing, in detail, how routine social

actions involving such machines are accomplished interactionally (Liesenfeld and Dingemanse, 2024; Porcheron et al., 2018).

We start from Garfinkel's (1967: 157) ethnomethodological conceptualization of "passing" as the "work of achieving the ascribed status" of, in this case, a human interlocutor. Garfinkel's (1967: 118–185) famous case study shows how Agnes, a transgender woman whose gender is under "chronic threat or open contradiction," uses various situated "passing devices" to protect her gender identity across a range of everyday and institutional interactions. Agnes' passing devices include euphemism, feigned ignorance, and other contingent strategies to "avoid any tests she thought she might fail" in everyday "passing occasions." The key point that Garfinkel (1967: 180) makes is that Agnes is a "practical methodologist" of "natural, normal female" social life whose practices do far more than conform to a set of dualistic gendered norms or suppress a fixed catalog of "tells." Indeed, binary gender "tests" based on definitional characteristics that ignore the situated performativity of gender can result in acts of misgendering (Pino and Edmonds, 2024) that can include persecuting cisgender people as trans (Joubin, 2024). Instead, Agnes learns to recognize and manipulate the "unavoidable, unnoticed texture of relevances" that embed "appearances-of-normal-sexuality" (Garfinkel, 1967: 183) in daily life.

This notion of "passing" presents a radically different challenge both to common interpretations of "passing the Turing Test". It neither aims to ascribe intelligence to machines nor does it, like the fictional *Voight-Kampff Test* of the eponymous bounty hunters in *Blade Runner*, aim to place suspects into untroubled categories of either "AI" (Suchman, 2023) or "human". Instead, in this article, we explore the practical and narrative potential of a "Conversational Action Test" (CAT) that explores the interactional work required to achieve conversational participation as constituted in specific, situated, interactional environments. Here the unit of analysis is not the "person" or the "AI." Rather, we focus on the mundane, interactional "passing occasions" within routine service calls, where callers and call-takers encounter one another within the limited roles and tasks involved in, for example, making a booking or enquiring about prices. In such highly constrained environments, "passing" as *human* is hardly the central challenge. Indeed, where we encounter an artificial agent "unannounced" as such, passing as human may still, though perhaps not for much longer, depend more on the basic assumptions or "trust conditions" that underpin a sequentially structured social interaction than on technical sophistication (Ivarsson and Lindwall, 2023; Relieu et al., 2020; Turowetz and Rawls, 2021). Participants may reasonably assume they are talking to a human simply by answering the phone and falling into the pervasive, mutual accountability of social interaction (Coulter, 1979). Our analysis, then, explores the interactional details of human-human service calls (e.g. to a doctor's surgery or a veterinary practice, or a university contact center), alongside a range of similarly task-constrained service calls performed by an AI conversational agent to human call-takers.

While we can categorize these calls, *a priori*, as "human-human," or "human-AI," such categorizations are neither the starting point nor the end goal of our analysis. Instead, we start with "detailed, concrete observations and descriptions of organizationally achieved social phenomena" in a routine service call (Garfinkel, 2021: 19; see also Eisenmann and Lynch, 2021). Turowetz and Rawls (2021) argue that Garfinkel's

focus on the lifeworld of marginalized identities with "at best, unstable routinization" (Garfinkel, 1967: 179) allows us to study the practical ethno-methods that members in human sociality use to "pass" or avoid contingent "tests we might fail." Examining a range of service calls where at least one caller, as Suchman (2023: 4) puts it, "travel[s] under the sign of AI," provides a rich opportunity for analytic observation. This approach also suggests a novel paradigm for developing evaluative tests of conversational AI based on empirical analysis of the "passing occasions" constituted through social situations.

Why test "interaction" rather than "intelligence"?

Interaction is far more explanatory and generative as an empirical material than reductive tests of ostensible intelligence. Most varieties of "Turing Test" use human judges to evaluate machine responses to text-based question-answer sequences as an operational test of "human-level intelligence" (Loebner, 2009), but often overlook the empirical material of interaction itself. Conversation analysts, by contrast, treat interactional resources and practices as their fundamental objects of study. CA has often studied the kinds of standardized question-answer sequences used in Turing Tests in a range of interactional settings. Such question/answer sequences usually structure common "interview activity types" (Levinson, 1979) that place routine, situated, interactional constraints on turn-by-turn talk. These patterns organize how participants solicit and produce accounts (Carlin, 2006; Potter and Hepburn, 2012), and an interactional perspective can explain *how* (not just *that*) such tests are "passed." For example, Weizenbaum's famous ELIZA bot exploits the interactional constraints of question-answer sequences by reversing pronoun pairs from "your" to "my" in each turn (Wallace, 2009). Critics who decry this kind of passing as algorithmic "trickery" rather than an ostensible test of "AI" (Harnad, 1992; Kurzweil and Kapor, 2009) often suggest making the test harder by, say, extending its length or topical range. However, this overall approach fundamentally treats "intelligence," operationalized by interaction, as somehow separable from the interactional structures and practices on which the test itself relies, risking "losing the phenomenon" (Eisenmann and Lynch, 2021) entirely.

By contrast, Collins (2018: 50–51) argues that a well-designed test should focus on the "quintessentially human activity" of *repair*: the ways participants deal with "problems of speaking, hearing and understanding" as they occur within social interaction (Jefferson, 1987; Schegloff et al., 1977: 361). Repair operates as a naturalistic, endogenous, "test" of mutual understanding by enabling coordinated joint action (Albert and de Ruiter, 2018). Contrast this with exogenous "tests" where human judges decide, post hoc, whether the participants' responses to test questions have matched their assumptions about "human intelligence." Given the universal availability of repair across languages and cultures (Dingemanse and Enfield, 2024), we can track, monitor, and re-establish mutual ongoing intersubjectivity in interaction if or when it seems to be breaking down. For example, one can initiate repair at any time by flagging a "trouble source" and can enact repair by providing a "trouble solution" before progressing the interaction. The speaker of the trouble source ("self") and a recipient ("other") can use a four-way matrix of repair actions that are "self-initiated self-repair," "self-initiated other-repair," "other-initiated self-repair,"

and "other-initiated other-repair." Repair thus functions as an infrastructure for intersubjectivity (Schegloff, 1992) between "self" and "other" because each party can initiate and resolve repair at any time. Rather than defining an operational test for the intelligence or subjectivity of one party to an interaction, repair endogenously constitutes each party's subjectivity as a special case of intersubjectivity through interaction.

Similarly, the embodied interactional order is often overlooked in operational tests of machine intelligence, and in computational linguistics more broadly. As Goodwin and Heritage (1990) point out in a discussion of Chomsky's (2002) disregard of linguistic "performance," informational theories of communication that exclude the "noisy" data of talk cannot deal with how language is used interactionally. Thus, Natural Language Processing (NLP) technologies tend to treat repair, disfluencies, hesitations, glottal cut-offs and other "miscommunication phenomena" as informational noise by filtering them out (Healey et al., 2018). Such embodied interactional resources are, therefore, mostly ignored (Purver et al., 2018), despite their fundamental importance for recognizing, forming, and ascribing *social actions* (Levinson, 2013). As Pütz and Esposito (2024) demonstrate in their study of interactions with LLM-based chatbots, where repair *does* occur, it is the humans that do most of the interactional work. In summary, rather than operationalizing tests for "artificial intelligence" through post hoc human judgments *about* interaction, the CAT proposes examining conversation itself as a material and locus for the observable, endogenously analyzable "embodiment of human sociality" (Schegloff, 2015).

Why a CAT? And what should it test for?

The structural organization of social action is remarkably stable over time and between settings when compared to the situated contingencies of language and meaning (Heritage, 2008). A CAT, then, might draw on the way CA studies social action in specific settings as constituted by sequences of "turn constructional units" (TCUs) that build and progress courses of action (e.g. requests, offers, invitations), where any single action can be achieved via multiple grammatical formats. For example, "requesting" may be achieved by interrogatives (e.g. "can I"; "do you"; "would you") in some situations, but also by declaratives (e.g. "that cake looks good") or narrative descriptions (e.g. "I've been getting terrible headaches lately") in others. Such actions are also often defeasibly and tacitly embedded within "pre-sequences" such as "my car is stalled" produced as a precursor for a request for a lift (see Stokoe et al., 2024), or produced through embodied resources such as gaze, head orientation, or gesture (e.g. a "can I have the bill" gesture in a restaurant). In all cases, it is the *action*—the offer or request—rather than the specific words or practices that implement the action that is consequential for what happens next (e.g. an acceptance, granting, or rejection). Our selection between—and recognition of—one another's choices between methods for initiating and responding to social actions are what constitutes the situated specificity of human sociality (Goodwin, 2000). In this sense, social action is central to human sociality and could motivate our tests and evaluations of conversational technology (Liesenfeld and Dingemanse, 2024) in terms of situational constitutiveness; that is, the "realness" or "artificiality" of the sociality they achieve.

This approach stands in stark contrast with methods of automatic NLP, where social action is conceptualized as abstract "user intent," rather than concretely constituted through turns and sequences of social interaction (Albert et al., 2019). Even state-of-the-art LLMs cannot reliably address the long-standing "pragmatics problem" (Cummins and De Ruiter, 2014) of mapping between words and social functions (Stokoe et al., 2024). NLP systems that model the regularities of semantic and lexical features still focus on *language*, rather than *action* (Housley et al., 2019), missing out on the pragmatic context that shapes the relevance of any utterance. By "context," here we refer to the turn-by-turn construction of the prospective and retrospective interpretability of actions and utterances rather than to a generic "bucket theory" of psychological or cultural context (Goodwin and Heritage, 1990). While technologists acknowledge that "context matters" for the sense of any utterance (e.g. Pearl, 2016), it is also often presumed that a task or setting (e.g. a specific type of service call) supplies "context" as a fixed variable (Stokoe et al., 2021; Stokoe and Richardson, 2023). Pragmatic context, on the other hand, is dynamically constructed by local modifications of, say, the organization of turn-taking (see Albert et al., 2019), multi-unit turn design (see Relieu, 2024), or patterns of non-lexical vocalizations, disfluencies, and hesitations (Lopez et al., 2022), and these practices are CA's central object of study.

A CAT of Google Duplex

Here we use CA to examine an instance of what Natale and Depounti (2024) describe as a "banal deception": *Google Duplex*. At its launch, journalists enthusiastically described how this telephone reservation and inquiry-making bot used "pauses and 'ums' to mimic a human" (Chen and Metz, 2019), and—within the limitations of its booking task—to interact "flawlessly" enough to "believe the hype" (Amadeo, 2018). These mirror later journalistic responses to the launch of ChatGPT and other LLMs in the early 2020s. In the analyses below, we focus on interactions initiated by *Duplex* in its publicly available recordings. Our observational focus is informed by related analyses of a wide range of pragmatically similar service calls drawn from the cumulative body of systematic research (including our own previous work) on social interaction in service calls. Building on these analyses, we outline procedures for conducting a putative CAT. We suggest the CAT as a practical method for creating situationally specific threshold criteria for the competences (including those of "AI"-labeled participants) associated with interaction in routine service calls. We then discuss how the procedures and criteria for a CAT may be adapted and replicated for drawing new empirical and conceptual axes for future comparative and applied studies in the field of conversational AI.

Data and methods

Some of CA's earliest findings document the structure of call-opening sequences (Sacks, 1995, pp. 3–32; Schegloff, 1968). Our analysis uses three data sets that are rich in these routine actions. First, we used the collections of "classic CA data" currently in circulation (Hoey and Raymond, 2022), for example, the Schegloff Media Archive (International Society for Conversation Analysis (ISCA), 2023), featuring hundreds of

call openings, appointment-requests, and other routine actions within a range of service call environments. Second, we used several large sets of between 100 and 3000 call recordings from our own previous studies of service calls to doctors' offices (Stokoe et al., 2016), university administration contact centers (Hoey and Stokoe, 2018), and veterinary surgeries (Stokoe et al., 2020). Our third data set consisting of a set of service call recordings featuring Google *Duplex* allowed us to compare actions in human-human service calls to related routine actions in *Duplex*-human calls.

We were able to access *Duplex* calls from publicly available recordings produced in Google's promotional material and technical documentation, although these data came with some analytic and ethical complexities. We first downloaded and transcribed all available Google *Duplex* calls using Jeffersonian transcription (Hepburn and Bolden, 2017), totaling five complete encounters and several smaller fragments (Leviathan and Matias, 2018). These calls seem to have been edited before publication, possibly for data privacy reasons. We assumed, a-priori, that these were all *Duplex*-human calls, although Chen and Metz (2019) revealed that Google uses human call-center workers for up to 25% of its *Google Assistant* app calls, while *Duplex* handles the rest. Where *Duplex* fails in these calls, the call is transferred to a human operator. One such recording published online by the *New York Times* (Chen and Metz, 2019), provides us with at least one like-for-like comparison between *Duplex* and its human counterpart. We selected calls in which the main purpose was closest to the Duplex calls (e.g. booking appointments for non-urgent services such as annual vaccinations). We used these calls as publicly available data, since they are published online, though we recognize that no explicit consent was given for this research purpose. Nor, for that matter, was consent for this use necessarily given by participants in the calls collected in CA's canon of "classic data," published long before contemporary norms of institutional ethical review. Nonetheless, the public, online availability of these data rendered them fair use for our research purposes. Participants in our corpus of 500 human-human service calls consented to us using these recordings for research purposes.

In the analyses below, we follow Schegloff (1987, 2009) by applying previous findings about specific interactional phenomena to new data and by taking a comparative approach. The range of interactional phenomena we focus on here were inductively derived from repeated reviewing and analysis of our data, informed by the wealth of existing CA findings about the structure of service calls (e.g. Flinkfeldt et al., 2021; Hoey and Stokoe, 2018; Lee, 2006, 2011; Schegloff, 1986; Stokoe et al., 2016, 2020; Whalen et al., 2002). We begin each analytic section by outlining an interactional practice identified in previous CA studies of human-human service calls, using examples to describe the interactional features that constitute the phenomenon. We then analyze *Duplex* calls featuring similar phenomena to see how the actions in question are recognized and accomplished (or not). We aim to show how a "baseline" analysis of routine interactions in a specific environment (here, service calls) can draw on the wealth of interactional research in similar settings to underpin a comparative analysis. A further aim is to also show how such analyses allow us to evaluate the ostensibly "artificial" sociality constituted by the actions of an AI participant. We should note here that our designation of "artificial" and "AI" here is made *a priori*, and is, in any case, not the point of this analytic exercise. Whatever our ontological commitments, our analyses only commit to

these *a priori* categories as a convenient starting point for analysis that focuses on methods and practices, not individuals, intelligences, or persons (artificial or otherwise).

Analysis

We present five sections of analysis. In the first two, we examine turn-component and sequential aspects of call openings, in which callers produce (a) first turns in the "reason-for-the-call" slot and (b) "second summonses," in which callers extend openings by redoing a summons before progressing to the reason-for-the-call. In three further sections, we examine features of trouble, perturbation, and repair in which callers (c) place and produce "um" and "ah" particles in the unfolding production of turns; (d) mark trouble; and (e) organize and respond to repair initiation. In each of the extracts below, some of which predate mobile and video telephony, we should note that all calls are audio-only. While this provides a somewhat restricted interactional environment where participants cannot see one another, talk is still rich with forms of phonetic embodiment available to both parties through prosodic and intonational variation. We also, therefore, offer some phonetic observations of *Duplex's* vocally embodied performance, based on acoustic and impressionistic approaches to comparable human-to-human calls. Together, these analytic approaches allow us to identify *Duplex's* capabilities and shortcomings and to reflect on their implications for testing the artificial (or otherwise) sociality of its routine actions.

Reason-for-the-call in service call openings

The first challenge for all participants in service calls, human or otherwise, is to conduct the situationally relevant organization of the call-opening sequence (Whalen and Zimmerman, 1987). The interactional features that constitute this routine include a summons/answer sequence, a greeting from the call-taker, and an official "place-self-identification" (e.g. a business name, Schegloff, 1986: 123). The call-taker usually speaks first, so the criterion for success in this routine is successfully moving from the call taker's answering the ringing phone to delivering the reason-for-the-call. This usually involves placing a service request in the "anchor position" (Schegloff, 1986): the structural slot in the opening where the caller may introduce the first topic. Reaching this point is criterial for a successful call opening because it demonstrates having achieved and progressed beyond mutual recognition of caller and call-taker's respective roles.

Extracts 1–3 show human-human calls to the vet (extract 1) and doctor's (2–3) receptions.

Extract 1: RC-jabs 2

```
01  REC: Dunnetts Vets.=Highuptown, Maggie speaking, how can I
02       ↓he:lp.
```

```
  03              (0.4)
→ 04  CALL: Hello there:.=um: I need t'make an appointment t'bring
  05        the cat in t'get its um: updated vacci↓na:tions.
```

Extract 2: GP-61

```
  01  REC: ↓Good afternoon, Limetown surgery, Tracy speaking?
  02             (0.3)
→ 03  CALL: .Hh oh:, good afternoon.=could I book (0.3) a flu jab please.
  04       =[for myself an' m'husband.
  05  REC: [Yes o'course.
```

Extract 3: GP-75

```
  01  REC: Good mornin'.=Limetown Surgery:,
  02             (0.5)
  03  CALL: #Ah# goo' mor'ing.
  04             (.)
→ 05  CALL: (I'm/Ahm) ringing t'make an appointment with the
  06       nurse please,=an' have my ears syringed.
```

In the three extracts below, *Duplex* (DUP) calls reception (REC) at a salon and two restaurants to make bookings. Each includes all the routine components of a service call opening, albeit with the identification components apparently redacted. *Duplex* first provides a responsive greeting (e.g. "H↑i::.") then requests a booking as the reason-for-the-call in the anchor position.

Extract 4: salon1 (http://bit.ly/duplex-salon1)

```
  01       ((Phone rings))
  02  REC: ((identification redacted)) h'llo how can I help you.
  03             (0.7)
→ 04  DUP: H↑i::. I'm calling to book a women's haircut for a
  05       cli:ent?
  06             (.)
  07  DUP: U:m.
  08             (.)
  09  DUP: >I'm looking for something< on May thi:rd?
```

Extract 5: booking_a_table-2 (http://bit.ly/duplex-table2)

```
  01       ((Phone rings))
  02  REC: ((identification redacted)) >>how may I help<< you?
  03             (0.9)
→ 04  DUP: He:y. I'::m calling to make reservation?
```

Extract 6: booking_a_table-1 (http://bit.ly/duplex-table1)

```
01        ((Phone rings))
02              (0.8)
03   REC: ((identification redacted)) >>Hi how may I help you<<?
04              (0.9)
→ 05   DUP: Hi::. U::m I'd like to >reserve a table< for Wednesday the::
06        ↓seventh.
```

Extract 7 is from our one recording of a call initiated by a human Google call-center worker. The same opening sequence is accomplished, but in this, the business self-identification (the restaurant name) is unredacted.

Extract 7: nytimes-restaurant_booking (http://bit.ly/duplex-nyt-1)

```
01   REC: Lao Thai Ki:tchen,
02              (1.5)
→ 03   GOO: Hello,
04              (0.2)
→ 05   GOO: I'm callin' make a reservation for a client
```

If we compare the human-human service and human-*Duplex* calls, we see similarly structured opening sequences containing the same turn components (e.g. greeting, request, etc.), which reflexively accomplish the mutually recognized interactional roles and actions of a "service call." In these types of sequences, then, based on an examination of routine procedures, a CAT would define a criterion for "passing" at a threshold for interactional competence that *caller reciprocates any greeting and moves on to the first topic in the next turn*.

Re-setting the call opening via a second summons

In some situations, of course, the routine turn components of call openings may be organized somewhat differently. As we have seen, in service calls, the summons of the ringing phone is usually reciprocated with a vocal response including various routine components (e.g. greetings, self-identifications etc.). Where the call-taker's vocal response is missing, previous interactional studies have identified the "second summons" as a method callers can use to deal with the absence of the vocal response. For example, if the caller does not hear the call-taker's responsive "hello," perhaps due to a technical problem, they may re-do their initial summons (i.e. the ringing of the phone), with a spoken, often upward intoned, re-doing of the summons turn, for example, "hello?" (Schegloff, 1968: 1088). Second summonses are also useful for dealing with other kinds of call-opening trouble. For example, Lee (2006) showed that Korean callers often do a second summons if they have not recognized the call-taker's voice, which can occasion a repeat response, providing the caller with another opportunity to identify the call-taker from their voice sample. In all cases, the second summons works by sequentially deleting

whatever the call-taker may have said in their initial summons response and making a re-doing of the response relevant next. A second summons is successfully achieved, then, when the call-taker re-does their summons response.

Extracts 8–9 provide examples of second summonses from a variety of human-human service call settings including calls to the police and to university admissions:

Extract 8: From Schegloff (1968)

```
01   D:   Police Desk (pause). Police Desk (pause) Hello, police desk
02        (longer pause). Hello.
→ 03 C:   Hello.
04   D:   Hello (pause). Police Desk?
```

Extract 9: CC-01

```
01   UNI: .mkhh Good afternoo:n, Browton University contact
02        centre.=Anne speakin'?
03             (1.1)
04   CALL: .tkhh Hello:.
05   CALL: .hhhh
06             (1.1)
07   CALL: H[ e l l]o:.
08   UNI:    [hello?]
09             (0.9)
10   CALL: Hell[o Anne
11   UNI:       [(he-    )
12   UNI: H[ello
13   CALL: [U:hm .tk I'm d-
14             (0.3)
15   UNI: [Mm? ]
16   CALL: [Hell]o:. I'm calling up on behalf of my dau:ghter.
17             (.)
18   CALL: Who is away at the moment.=but she's ↑just had her:- (.)
19        aye level ((final school examination)) results.
```

In extract 8, the Police Desk dispatcher does an official self-identification as a first response, then the caller does a second summons in line 03, occasioning a full repeat of the dispatcher's first summons-response turn. Note that the second summons here achieves a "reset" of the call when the dispatcher then "re-starts" with a full repeat of the official summons response and institutional identification "Hello (pause). Police desk?" in line 04. In extract 9, the caller is a parent calling university admissions on behalf of their child. The second summonses here deal with troubles of overlapping talk. The caller's second summons in line 07 comes after a series of delays (lines 03–06) that occasion an overlapped response (line 08). The caller then re-does a second summons adding the

call-taker's name "Anne" (line 10), once again in overlap. This time the call-taker duly re-does their summons response (lines 11–12) sufficiently in the clear to facilitate progress to the first topic at line 16, effectively re-starting the call-opening sequence.

Extracts 10–12 show how *Duplex* deals with trouble or deviations from the routine structure of service call openings using a second summons to accomplish a "reset" in the opening sequence.

Extract 10: booking_a_table-3 (http://bit.ly/duplex-table3)

```
01        ((Phone rings))
02   REC: ((business name redacted)) good evening,
03              (0.7)
→ 04   DUP: Hello:?
05              (0.8)
06   REC: Hello:,
07              (0.7)
08   DUP: HI::.
09              (.)
10   DUP: U::m I':d like to reserve a table for F:riday the
11        ↓thi:rd.
```

Extract 11: asking_opening_hours (http://bit.ly/duplex-hours)

```
01        ((Phone rings))
02   REC: ((business name redacted)) ( ) how can I he:lp you.
03              (0.8)
→ 04   DUP: Hello:?
05              (0.6)
06   REC: Hell↑o:. >what's up man,<
07              (0.7)
08   DUP: He::y.
09              (.)
10   DUP: U:::m I wanted to know what are your hours for
09        ↓today.h
```

Extract 12: duplex_restaurant_call-nytimes (http://bit.ly/duplex-nyt-2)

```
01   REC: >>Hello Bowl'd Solano <how may I<< help you?
02              (0.8)
→ 03   DUP: Hello?
04              (1.9)
05   REC: Hello,
06              (0.8)
07   DUP: Hi:: I'm calling to make a reservation?
```

In each case, *Duplex* issues a second summons following the call-takers' first response. This second summons has a rising pitch contour—common in standalone first greetings in English (Kaimaki, 2011). In the calls above, following Duplex's second summons, the call-taker duly re-issues a response, sequentially re-setting the call opening. In each case, in the following turn, *Duplex* proceeds to the first topic, as in the straightforward call openings in Extracts 1–7.

Both *Duplex*'s and human callers' second summonses above clearly create an opportunity to re-start the call-opening sequence, so a CAT might treat the reset of the call following a second summons as a criterion for successful service call interactions.

Anchor position uh(m)s

Duplex's developers note that where a response may be expected with no delay, or when dealing with complex activities that may incur what Leviathan and Matias (2018) call "processing delays," *Duplex* may interject a "speech disfluency" or a sound stretch that "masks" such delays. However, as Schegloff (2010) points out, these utterances have a wide range of systematic sequential positions, functions, and production characteristics far beyond covering for delays. For example, in a call opening sequence, callers routinely produce an "um," "uh," or "ah" (all of which we combine here as "uh(m)") just before the reason-for-the-call in "first topic" slot (Schegloff, 1986). This is a different phenomenon from the type of uh(m) that often occurs when participants encounter troubles of speaking or understanding (e.g. Jefferson, 1974). Callers can also produce a first topic without doing a turn-initial uh(m); however, pre-anchor position uh(m)s can project the reason for the call or some form of intersection rather than trouble, as suggested by the way they also occur when the anchor position is "displaced" by some other business (Schegloff, 2010).

Extracts 13–15 below are taken from human-to-human service calls to GP offices, vets, and police dispatchers. In each case, the caller produces this specific type of anchor position uh(m).

Extract 13: GP-84

```
01   REC: ↑Good morning Limetown ↑surgery,
02              (0.3)
03   CALL: ↑Good morning,
04              (0.3)
→ 05 CALL: ↑U:hm, I need to make an appointment for Brian Tristram Sadler.
06        =with Doctor Long please
```

Extract 14: RC-Vaccine 32

```
01   VET: Good afterno:on, Johnson Veterinary Centre, Joan
02        speakin'?=↑how c'n I ↓he:lp.
→ 03 CALL: .pt Oh hello Joan,=uhm >I've got an appointment< booked
04        for <Spar:ky Jackson t'see:-.h for a nurse clinic.=at
05        six tonight.=it's f'r socialisation, [.hhh
06   VET:                                      [Oka:y?
```

Extract 15: (from Schegloff, 2010, pp. 151, #17)

```
  01   Dis: P'lice Desk,
→ 02   Cal: Uh, could you uh go to uh leven twenny five Broadway,
  03   Opr: Yes, please,
  04   Dis: We're talking operator, go ahead sir,
  05   Cal: Uh could you go to leven twenny five Broadway
  06        Apartment five, and uh tell the lady that answers
  07        the door that uh (1.4) this is uh her husband
  08        (uh)/(en) (0.5) he's been uh,(0.2) I've been picked
  09        up by the state police, (0.2) no tail lights on the
  10        truck, (1.5) and uh (0.8) be home late. Wouldja-couldja
  11        give 'er that message?
  12   Dis: Where are you now.
```

In Extracts 13 and 14, the caller reciprocates the greeting before doing an uh(m) and moving on to provide the first topic in anchor position. Schegloff (2010) uses extract 15 to demonstrate the relevance of an anchor position "Uh" (in line 02), where the caller begins to ask for help and give an address. After the operator interjects and the dispatcher explains the interjection, note how the caller re-starts his request for emergency help (line 05). He repeats the "Uh" in anchor position but deletes the other two "uhs" ("could you uh go to uh") in his re-doing of his first topic turn, suggesting that *only* the anchor position uh(m) has some kind of persistent interactional relevance.

In Extracts 11 and 12, above, and in extract 16 below, *Duplex* produces an uh(m), or a sound stretch that sounds like an uh(m), just before introducing the reason-for-the-call.

Extract 16: additional_fragments (http://bit.ly/duplex-fragments)

```
→ 02   DUP: H↑i:: u:::m I would like t'reserve a table: for
  03        May twenny ↓fifth.
```

Despite the claims of the developers to be masking processing delays, the placement of these uh(m)s does not appear arbitrary. These are slotted into the anchor position when the opening sequence is extended in various ways. For example, in Extracts 10 and 11 above, and in extract 17 below, *Duplex* extends the greeting sequence by using a second summons to reset the call. In these cases, *Duplex* still produces an uh(m) in anchor position before introducing the first topic.

Extract 17: duplex_restaurant_call-nytimes (http://bit.ly/duplex-nyt-2)

```
  01   REC: >>Hello Bowl'd Solano <how may I<< help you?
  02              (0.8)
  03   DUP: Hello?
  04              (1.9)
  05   REC: Hello,
```

```
06              (0.8)
07   DUP: Hi::.=I'm calling to make a reservation? I':m Google's
08        automated booking servi:ce? so I':ll record the call?
09              (0.7)
10   DUP: A::m (0.3) could I book a table for Tuesday the
11        twenny first?
```

Note that here in lines 7 and 8, *Duplex* starts with an announcement about the "reason for the call" ("I'm calling to make a reservation?"), but without actually producing the reservation request. This turn functions as a kind of "pre-request" forming part of a standardized service announcement that the call is from an automated booking service and is being recorded. These types of pro-forma "recording for training and monitoring purposes" announcements are, typically, separate from the "business" of the call. Indeed, once the pro-forma announcement is delivered, *Duplex* does an anchor position "A::m" just before the first topic in line 10, suggesting that this uh(m) tracks the anchor position, rather than simply being placed after the greeting sequence automatically.

Whatever the interactional consequences of this phenomenon, *Duplex's* anchor position uh(m)s successfully occasion a re-doing of the call-taker's response and, as such, they achieve this interactional practice.

Other-initiated self-repair

One striking feature of *Duplex's* calls is that, in few instances, its calls involve the use of other-initiated self-repair (i.e. where "other"—the recipient of the trouble-source turn—flags the problem, then allows "self"—the speaker of the trouble source—to solve it). In human-human service calls such as in extracts 18-20, below, this kind of repair operation often occurs when it is especially important that participants achieve and secure a shared understanding of times, dates, and other consequential details.

Extract 18: GP-14

```
01   D: Good morning Surgery: Trish speaking?
02              (1.3)
02   C: Hello have you got an appointment for Fri:day:
03      (.) afternoon or teatime please.
04   D: This Friday,
05              (1.0)
06   C: Yeah.
```

Extract 19: GP-28

```
01   C: Anything with De- Doctor De Meyer at all?=
02   D: =(( inaudible, phone clattering )) lost me. (.) just a moment,
03      (12.4)
04   D: No I've not got anything with him at the moment,
```

```
   05      (2.8)
   06   C: Ri::ght?
   07      (1.9)
   08   C: E::hm, Doctor Keeg[an?
   09   D:                    [SIXteenth of October Doctor De Meyer,
   10      (0.9)
→  11   C: >(Which do- ah-)< sorry?
   12   D: SIXteenth of October.=
   13   C: That suits me perfect.
```

Extract 20: (from Heritage and Clayman, 2010: 75)

```
   01   911: Mid-City emergency.
   02   Clr: .hhh Yes. (.) would you th'polic:e (.) please to:
   03        twenty three forty four James North .hhh downstairs.
   04   911: Whatsa problem ma'am.=
   05   Clr: =U:::h (0.1) I just went by there and my son lives
   06        there an' his wife an:: thuh family,=
   07   911: =Uh huh,=
   08   Clr: =An' uh (.) there's some kids throwin' knives at
   09        their house.
→  10   911: Knives?!
   11   Clr: ^Yeah!
```

Note that there is a variety of forms of other-initiation that we see from the participants in these three cases. In extract 18, line 04, we see the call-taker (D) use a partial repeat of the prior turn as an "understanding check." "This Friday," with the stress on the proximal demonstrative pronoun "this," disambiguates the caller's prior, more vague reference to "Friday afternoon or teatime." This "understanding check" form of other-initiation does most of the work of the repair, since the recipient thereby explicitly demonstrates their understanding of the prior turn. The trouble-source speaker may then simply do a token confirmation such as the caller's "yeah" in line 06.

In extract 19, line 11, we see the caller (C) use a less specific "open class" repair initiator "sorry?" (Drew, 1997) in their call to the doctor's surgery. This type of repair initiator does not specify the trouble source, leaving it up to the trouble-source speaker to identify and resolve the problem. Here the call-taker (D) treats the problem as a mishearing of the date by repeating that component of their prior turn "sixteenth of October" in line 12, which the caller duly accepts. This form of repair initiation (often done with "what?" or "huh?"), leaves most of the work of identifying and repairing the trouble to the speaker. It is very common in situations (as in this case), where overlapping talk or some other sound coincident with the prior turn may have occasioned a mishearing (Dingemanse et al., 2014).

In extract 20, we see another example of the partial repeat and "understanding check" form of repair initiation. Here the 911 call-taker does an emphatic partial repeat of the prior turn "Knives?!." This prosodically marked re-doing of one word from of the prior turn "there's some kids throwin' knives at their house" does what Wilkinson and Kitzinger

(2006) call a "performance of surprise," highlighting the unexpected or extreme nature of the report. This case shows how repair procedures ostensibly used for solving problems of speaking, hearing, and understanding may also implicate broader issues such as inappropriateness or deviation from social norms.

In Extracts 21–23, we also see *Duplex* participating (as trouble-source speaker, or "self") in several instances of *other-initiated self-repair*. Note that Extracts 21 and 22 are from call fragments, so they cannot be analyzed in any wider sequential context.

Extract 21: additional_fragments (http://bit.ly/duplex-fragments)

```
((Recording fragment starts mid-call))
DUP: H↑i:: u:::m I would like t'reserve a table: for May
     twenny ↓fifth.
          (0.7)
REC: Sorr↑y ↓what ↑da:y?
          (0.9)
DUP: For Fri:day: a:::h May twenty fi::fth?
```

Extract 22: From "additional_fragments" (http://bit.ly/duplex-fragments2)

```
DUP: The:: phone number is (0.3) uh::m si:x oh:: se:ve:n
          (0.5)
REC: >Wait< (.) >>wait wait<< can you start ove::r?
          (1.5)
DUP: The:: number is (.) si:x oh:: se:ve:n
          (0.4)
REC: Uhuh,
          (1.1)
DUP: Two two-
     ((recording is cut off))
```

Extract 23: From "booking_a_table-1" (http://bit.ly/duplex-table1)

```
DUP: Hi::. U::m I'd like to >reserve a table< for Wednesday the::
     ↓seventh.
          (1.8)
REC: Fo::::r seven people?
          (1.2)
DUP: U::m (.) it's for four people.
          (0.8)
REC: Four peo↑ple ↓when:::,
          (1.7)
REC: [Today? tonight?]
DUP: [U::m       ne]xt Wednesda:y? a:t six pee em.
```

In extract 21, after the call-taker initiates repair at line 05 with "sorry what day?," *Duplex* provides the repair solution, inserting the day "Friday" as well as repeating the relevant part of the trouble-source turn "May twenty fifth." This repair treats the trouble either as an issue of which day of the week the reservation falls on, or as a trouble of hearing/understanding the day and date altogether.

Extract 22 starts as *Duplex* is dictating a phone number when the call-taker initiates repair by asking *Duplex* to "start over." *Duplex* duly re-starts the dictation, and this time the call-taker displays uptake and alignment by doing a "continuer" (Goodwin, 1986) "Uhuh" at line 33 after the area code, interspersed between Duplex's ongoing dictation.

In extract 23, the call-taker's turn at line 08 "Fo::::r seven people?" initiates repair with an "understanding check," offering a candidate hearing of *Duplex's* prior request for a reservation on "Wednesday the:: ↓seventh." (lines 05 and 06). The call-taker's candidate hearing in line 08 turns out to have been a mishearing since in line 12, the call-taker goes on to ask about the reservation date: "Four people . . . when," then, after a pause, "Today? Tonight?." If the call taker had heard Duplex's prior turn as "Wednesday the seventh," the issue of the date would have been clear already. *Duplex's* repair in response to the understanding check "U:::m it's for four people" unproblematically treats the call-taker's understanding check "for seven people?" as an incorrect guess as to how many people to reserve a table for, rather than as a mishearing of Duplex's initial request about a specific day/date (Wednesday the seventh). They deal with the outstanding issue of when the reservation is for in subsequent turns.

In terms of sequential structure, these three examples all demonstrate the successful accomplishment of other-initiated self-repair because following the repair procedure, both caller and call-taker proceed with the task at hand. However, *Duplex's* responses in Extracts 21–23 do not unambiguously display as specific an orientation to the trouble source as we saw in the human-human examples in extracts 18–20. For example, in extract 21, the call taker flags up the *day*, but not necessarily the *date* as the trouble source, but *Duplex*'s response in the following turn includes both the day (Friday) *and* re-does the date, disattending to the specificity of the repair initiation "sorry what day?." Similarly, in extract 22 where the call-taker asks *Duplex* to "start over" when giving a phone number (line 29) *Duplex* re-does a fully sentential turn prefaced with, "the number is . . ." rather than responding to the precision of the repair initiation to "start over," that is, specifically re-dictating the number, rather than re-doing the entire turn. Finally, in extract 23, *Duplex's* repair in line 10 "um it's for four people" goes along with the call-taker's misunderstanding that the prior request related to numbers of people, without addressing their prior mishearing of "Wednesday the seventh."

So, while *Duplex's* responses to other-initiations of repair in these data meet the basic sequential criterion for accomplishing repair (i.e. getting the repair done and moving on), our final analysis in this section suggests a certain lack of sensitivity, on *Duplex's* part, to the precision of other-initiations of repair to locate and help to swiftly resolve interactional trouble. In the following section, we discuss how our analyses, starting with human-human service calls, show how we might develop a CAT that evaluates the performance of callers or call-takers (artificial or not) in terms of their participation in situated forms of sociality.

Discussion

The aim of this article was to examine how a conversational voice agent interacts on the phone with naïve human interlocutors in service encounters to achieve a form of "banal deception" (Natale and Depounti, 2024). We evaluated *Duplex*'s turns in relation to conversation analytic research into the structure of call openings, second summonses, uh(m)s that precede a reason for the call, and other-initiated self-repair. We evaluated *Duplex*'s achievement of these practices against human-to-human calls, following Schegloff's (2009) guidelines for comparative CA that require analysts to describe the interactional features that constitute a practice, to propose criteria to test its achievement, to discuss how the practice may transfer to other interactional contexts. Our analysis showed that *Duplex's* actions largely achieved these practices in terms of our basic procedural/sequential criteria. In the following section, we discuss how each practice "passes" as conversationally competent and ask what we can infer from observing the degree of specificity with which *Duplex* responds to other-initiations repair. We consider the broader implications of using CA in situations that resemble the fictional "Voight-Kampff Test" for artificial sociality. Finally, we propose some aims and procedures for developing a form of "CAT" capable of evaluating sociality in specific interactional situations.

"Passing devices" maintain artificial sociality

Our analysis highlighted several methods that *Duplex* used to progress through a potentially tricky interaction. First, the opening phase of a service call is a highly routinized site for institutional talk, where contributions from each party fit into a set of mutually expectable sequential "slots" (Drew and Heritage, 1992), although there may still be significant variations. As our human-human data reveal, greetings can vary with time of day (good morning/evening); and may include names and organizational self-identifications that can be more generic or more specified (e.g. "surgery" vs "Limetown Surgery"). *Duplex's* practices for moving from call openings into the reason-for-the-call are clearly robust enough to manage these variations. However, even though *Duplex's* practices meet our criteria for achieving this call opening structure, they may not make use of all the interactional resources available. Minor variations in the position and composition of turns provide participants with a range of resources for accomplishing their respective, situated identities as they move on to the first topic of the call (Psathas, 1999). For example, in extract 9, the caller's long gaps, pauses and disfluencies display hesitation or delicacy in formulating her situated identity as the "mother of the official caller." *Duplex's* relatively crude use of second summonses in Extracts 10–12, on the other hand, simply reset the opening sequence, shunting the call toward first topic. We can thus see this use of second summons as one of several "passing devices" (Garfinkel, 1967): methods for moving through a stretch of interaction where there is a threat of exposing possible "incompetence." This method is very similar to how *Lenny*[3] (a telephone "spam trap" bot that simply reads out—with "a soft and slow Australian accent in the manner of an elderly man" (Oberhaus, 2018)—a set of 16 carefully scripted, pre-recorded turns to fool telemarketers into wasting their time, see Relieu, 2024; Sahin et al., 2017) occasionally

reports trouble on the line: "hello? are you there?," often resulting in re-setting, and sustaining the ongoing interaction.

Some passing devices effectively mimic the way people manage and mark trouble in ongoing talk through delays, disfluencies, and hesitations. The uh(m)s of this sort were enthusiastically applauded by the crowd during a demonstration of *Duplex* at the Google IO 2018 keynote (Google Developers, 2018), as well as in media reports that celebrated *Duplex's* "authentic" use of speech disfluencies. Indeed, our analysis showed that *Duplex* sometimes positions uh(m)s in ways that account for their placement (e.g. in overlap resolution, or in call openings just prior to the reason-for-the-call) and build toward a target action such as requesting a reservation. However, though we lack space to reproduce them here, our wider analyses of *Duplex* calls also found uh(m)s that seemed phonetically and procedurally unfitted to their sequential environments. Perhaps these were masking non-interactional "processing delays," as the developers claimed (Leviathan and Matias, 2018), rather than being positioned in relation to the unfolding action. Similarly, in "mystery shopper" calls described by Stokoe et al. (2020), mystery shopper callers simulating clients to test the phone services of a vet's surgery simply have different issues at stake from genuine pet owners, and thus use different interactional patterns. For example, while real pet owners answered the receptionists' questions about their pets fluently, mystery shoppers tended to delay, defer, or respond disfluently. Given the way that humans struggle to simulate the behaviors of other humans, even in task-specific contexts such as service calls, we might expect this to remain a long-term challenge for artificial sociality.

Finally, while *Duplex's* involvement in other-initiated self-repair is successful, it is also ambiguous since its responses do not always target the specific trouble source cited in the repair initiation[1]. These passing devices may help to smooth the path toward a successful service call closing, but the way *Duplex* uses them to "bypass" trouble may obviate valuable interactional resources humans use to recognize and deal with *miscommunication* (Healey et al., 2018; Purver et al., 2018). Indeed, we may depend on the specificity of our abilities to recognize and manage interactional trouble to secure shared understanding and intersubjectivity (Albert and de Ruiter, 2018; Schegloff, 1992; Sidnell, 2014). Where artificial forms of sociality *evade* repair using a passing device, they may miss an essential, if difficult, step toward understanding and dealing with more unpredictable and complex interactions.

The implications of AI for CA

One outcome of our analysis is to add a new analytical frame to CA, which has included, from the outset, a burgeoning set of studies framed as "institutional talk" in a wide range of settings including helplines, healthcare, and service interactions (e.g. Drew and Heritage, 1992). The structure of talk in these situations is studied in relation to the institutional constraints we can observe on the putatively ubiquitous frame of "everyday talk," which is understood to encompass a relatively unconstrained range of interactional practices (Hester and Francis, 2001). Each situated form of human sociality, described in terms of the constraints on "institutional talk," creates a "unique 'fingerprint' for each kind of institutional interaction" (Heritage, 1997: 225), providing the basis for informative comparative and

evaluative analysis. For example, Stokoe (2013) shows how, even when domain experts set out to simulate an interaction, such as police interview trainers in a role-play, they tend to talk in ways that do not correspond with recordings of real interviews (see also Atkins, 2019; Stokoe et al., 2020). Similarly, CA studies of "atypical interaction" (Antaki and Wilkinson, 2012; Wilkinson et al., 2020) involving disabled people, often in institutional settings, increasingly focus on how people manage constraints on normative interactional patterns rather than on the communication impairments or medical diagnoses of individuals (Bottema-Beutel et al., 2021; Maynard and Turowetz, 2022). In this vein, studies of interactions involving artificial agents may require new analytic frames that can evaluate, for example, conversation design, voice user experience design, and agent design etc. in relation to the specific "fingerprint" of practices and interactional competences that constitute a growing range of contingent, situated, socialities (cf. Porcheron et al., 2018). Such frames will need ongoing revision as interactional studies of artificial sociality extend further beyond task-specific domains of the HCI lab and become an increasingly ubiquitous part of everyday life (Mlynář et al., 2024).

Another implication of our analyses is to show how some interactional phenomena can be amenable to both automated and conversation analytic forms of discovery. The AI methods underpinning *Duplex* bear comparison, in some ways, with CA in that they are strongly data-driven and use observations as a basis for theorizing about phenomena that, as Sacks (1984) puts it (p. 25) "can find things that we could not, by imagination, assert were there." *Duplex's* use of anchor position uh(m)s is a good example of this kind of phenomenon. The machine learning methods that inform some of *Duplex's* behaviors may have "discovered" this little-known pattern of behavior, bottom-up, by deriving statistical regularities from processing large numbers of recorded service calls. *Duplex's* competent use of this practice therefore addresses some long-standing debates about whether, and how, some CA findings may be amenable to statistical and computational analysis (Button, 1990; Kendrick, 2017; Schegloff, 1993; Stivers, 2015). Although the interactional consequences of anchor position uh(m)s are still unknown, future studies that use AI in this way may identify related patterns in large volumes of data, opening up the possibility of using detailed CA studies to discover their situated interactional relevance (Steensig and Heinemann, 2015).

A research trajectory for a CAT?

When *Duplex's* practices and actions pass, and its service calls progress sequentially, this does not equate to *Duplex* itself "passing a Turing Test" in the vernacular sense of "passing as human." We call the method used in this article the "CAT" to focus, instead, on the actions and practices that comprise conversational competence and membership within specific interactional situations. We started by examining service calls involving Google's *Duplex* along with a wealth of data and findings from prior CA studies of similar interactional settings for comparative analysis. This enabled us to identify, describe, illustrate, and evaluate practices associated with the conduct of competent service encounters and their mutually acknowledged interactional roles. This analytic procedure comprises a test specifically configured for a particular interactional situation. This process may be repeated to design new CATs to evaluate how any agents (human or machine) achieve

conversational competence and membership across a range of interactional situations. In this way, the CAT can inform the design and evaluation of AI and voice technologies and may lead to new research questions for CA studies.

To design a CAT for a specific interactional situation, we suggest the following:

1. Specify an interactional setting underpinned by CA research.
2. Gather data featuring candidate actions and practices involving a "tested" party.
3. Gather data of normatively achieved practices in a similar, naturally occurring setting.
4. Transcribe and analyze data from both using standard CA methods.
5. Identify evident candidate practices for a comparative CA analysis (Schegloff, 2009).
   (a) State a clear understanding of the target phenomenon or practice.
   (b) Identify situationally specific and observable criteria for recognizing it.
   (c) Describe how this phenomenon has been examined and analyzed previously.
   (d) Compare its use between these environments and discuss any differences.
6. Ask if the tested party uses practices competently and is treated as a member.
7. Identify problems or observations that may feed into future design processes.

Having proposed this procedure for developing CATs, we conclude with a discussion of the implications for the design and interpretation of such tests more broadly.

The CAT evaluates actions, not agents

The CAT evaluates *social actions* rather than purported "intelligence"—artificial or otherwise—let alone ascribing the category of human or machine. Even if it were straightforward to ascribe humanness and evaluate intelligence, this common interpretation of the "Turing Test" has already been passed many times by simple chat bots (see Wallace, 2009) during the annual Loebner Prize competition (Loebner, 2009) with little impact beyond temporary sensationalist news coverage. Passing this kind of operationalized test of "human intelligence" often turns out to be trivial in both senses of being easy and being inconsequential. Some researchers have therefore advocated raising the bar for what might be considered intelligent up to and including being indiscriminable from a human (Harnad, 1992), or even exceeding human capabilities (Schweizer, 1998). Other proposals suggest extending the time allocated, stipulating the expertise of the judges, or enhancing the complexity or generality of the test (Kurzweil and Kapor, 2009). However, a harder operational test would not necessarily be any more explanatory about *how*, precisely, the test has been passed, nor what "indiscriminable" may mean in terms of how such judgments are made. Rather than refining operational tests that aim to ascribe human intelligence, the CAT aims to describe and then evaluate the pragmatics of situated human sociality. It describes criteria for evaluating the detailed interactional procedures that constitute each action, at each "passing opportunity." The analytic procedure of the CAT, using CA, can also provide thorough explanations about precisely how an action "passes" in each specific circumstance under scrutiny. For example, Sahin et al. (2017) use CA to show how *Lenny*'s call opening turns are designed to maximize coherence and agreement, to report "trouble on the line," and use misplacement markers such

as "by the way" to account for any incoherence with the caller's prior turn. *Lenny's* simple recordings are effective without using speech recognition, AI, or any NLP technology aside from playing pre-recorded turns when it detects that the caller has stopped speaking. Passing as human, then, which *Lenny* achieves with remarkable consistency, may rely more on the normative expectations that constitute the social situation, than on sophisticated AI systems.

The consequences of passing a conventional Turing Test have often focused on mediagenic scare stories of AI or robots "taking over" (Whitby and Oliver, 2000). In the case of *Duplex*, its first demonstration at the 2018 Google IO conference (Google Developers, 2018) did raise serious ethical questions about whether an AI should masquerade as human in public life (O'Leary, 2019). Similarly, today's AI-driven social bots are often convincing enough to influence commercial and political choices by emulating social media users, so there is an increasing demand for research into methods for categorizing agents as human or artificial (e.g. Ferrara et al., 2016). The arms race between AI developers and AI-detection measures will drive the sophistication of such systems, but not necessarily explain or ameliorate the consequences of their social actions.

A CA-informed approach such as the CAT, however, which focuses on the analysis of social actions, can achieve far more than simply ascribing the category "human" or "non-human." It can also show how such categories are used as resources in the production of social actions. For example, Housley et al. (2017) focus on the actions of social media users to show how discursive formulations of membership categories in social media posts can ignite antagonistic readings and responses and open up the potential for spreading false or malicious information. Thus, epithets like "bot" and "troll" are now used as terms of abuse on social media (Ruck et al., 2019), often aimed at users accused of repeating provocative or propagandistic talking points. These categories are harnessed as resources for social action (i.e. doing insulting), rather than working primarily as technical or ontological ascriptions. In terms of social consequences, whether the agent of an utterance is human or not may matter far less than how their utterances are implicated in a specific interactional situation.

Conclusion

The conceit of the Voigt-Kampff test in *Blade Runner* is to ask whether, and how, we define "humanness." The moral confusion of the protagonist *Deckard*, who falls in love with an android, shows how our intuitions, as well as more technical and conceptual operational definitions of humanness, may be fundamentally flawed. This focus on judging participants as either human or non-human by operationalizing interaction is a long-standing, though mediagenic, category error. Garfinkel's work on "trust conditions" showed how, turn-by-turn, interaction works as a "proving ground" for the micro-social structures and mutual expectancies that constitute human sociality. The categorical status of an interlocutor as "human" or "machine" is (still) rarely in question, whereas in everyday talk, the precise "fittedness" and the reciprocity of the design of each response to the previous action is, with each turn, immediately under scrutiny—as summed up in the conversation analytic dictum "why that now?" (Sacks et al., 1974: 241). Humanness, intelligence, and the artificiality (or otherwise) of sociality is not based on the inherent

properties of interlocutors but must be ongoingly constituted in and through action. With this proviso, we propose the CAT as a practical method for evaluating and understanding the coming wave of conversational AI through its constitutive involvement in forms of sociality. As Sacks (1995: 536) reminds us, "anthropomorphizing humans" is only an analytic convenience. For *Deckard*, in the end, "the electric things have their lives too." What matters is social action and how we conduct our social relationships in and through the technology of talk.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Saul Albert https://orcid.org/0000-0003-1043-1237
William Housley https://orcid.org/0000-0003-1568-9093
Elizabeth Stokoe https://orcid.org/0000-0002-7353-4121

Notes

1. See https://www.youtube.com/watch?v=ogfYd705cRs.
2. See, for example, the demos at https://www.gridspace.com/.
3. An archive of over 600 recordings of *Lenny's* conversations with spammers, phone scams, and telemarketers can be found at http://bit.ly/lennyarchive.

References

Adams R (2024) Researchers fool university markers with AI-generated exam papers. *The Guardian*, 26 June. Available at: https://www.theguardian.com/education/article/2024/jun/26/researchers-fool-university-markers-with-ai-generated-exam-papers (accessed 28 June 2024).

Albert S and de Ruiter JP (2018) Repair: the interface between interaction and cognition. *Topics in Cognitive Science* 10(2): 279–313.

Albert S, Housley W and Stokoe E (2019) In case of emergency, order pizza: an urgent case of action formation and recognition. In: *Proceedings of the 1st international conference on conversational user interfaces*, Dublin, 22–23 August, pp. 1–2. New York: Association for Computing Machinery.

Amadeo R (2018) Talking to Google Duplex: Google's human-like phone AI feels revolutionary. *Ars Technica*, 27 June. Available at: https://arstechnica.com/gadgets/2018/06/google-duplex-is-calling-we-talk-to-the-revolutionary-but-limited-phone-ai/ (accessed 27 June 2018).

Antaki C and Wilkinson R (2012) Conversation analysis and the study of atypical populations. In: Sidnell J and Stivers T (eds) *The Handbook of Conversation Analysis*. Hoboken, NJ: John Wiley, pp. 533–550.

Atkins S (2019) Assessing health professionals' communication through role-play: an interactional analysis of simulated versus actual general practice consultations. *Discourse Studies* 21(2): 109–134.

Bonifacic I (2022) Google is shutting down Duplex on the web. *Engadget*, 5 December. Available at: https://www.engadget.com/google-duplex-on-the-web-shutdown-announced-225937564.html (accessed 4 December 2022).

Bottema-Beutel K, Kapp SK, Lester JN, et al. (2021) Avoiding Ableist language: suggestions for autism researchers. *Autism in Adulthood* 3(1): 18–29.

Button G (1990) Going up a blind alley: conflating conversation analysis and computational modelling. In: Luff P, Gilbert N and Frolich D (eds) *Computers and Conversation*. Cambridge, MA: Academic Press, pp. 67–90.

Carlin AP (2006) Observations on features of a research interview. *Ciências Sociais Unisinos* 42(3): 177–188.

Chen BX and Metz C (2019) Google's Duplex uses A.I. to mimic humans (sometimes). *The New York Times*, 22 May. Available at: https://www.nytimes.com/2019/05/22/technology/personaltech/ai-google-duplex.html (accessed 23 May 2019).

Chomsky N (2002) *Syntactic Structures*. 2nd ed. Berlin: de Gruyter Mouton.

Collins H (2018) *Artifictional Intelligence: Against Humanity's Surrender to Computers*. London: Polity Press.

Coulter J (1979) The normative accountability of human action. In: Coulter J (ed.) *The Social Construction of Mind: Studies in Ethnomethodology and Linguistic Philosophy*. London; Basingstoke: The Macmillan Press, pp. 9–34.

Cummins C and de Ruiter JP (2014) Computational approaches to the pragmatics problem. *Language and Linguistics Compass* 8(4): 133–143.

Dingemanse M and Enfield NJ (2024) Interactive repair and the foundations of language. *Trends in Cognitive Sciences* 28(1): 30–42.

Dingemanse M, Blythe J and Dirksmeyer T (2014) Formats for other-initiation of repair across languages: an exercise in pragmatic typology. *Studies in Language* 38(1): 5–43.

Drew P (1997) "Open" class repair initiators in response to sequential sources of troubles in conversation. *Journal of Pragmatics* 28(1): 69–101.

Drew P and Heritage J (1992) *Talk at Work: Interaction in Institutional Settings*. Cambridge: Cambridge University Press.

Dwoskin E (2019) "I'm Google's automated booking service." Why Duplex is now introducing itself as a robot assistant. *Washington Post*, 27 June. Available at: https://www.washingtonpost.com/technology/2018/06/27/heres-why-googles-new-ai-assistant-tells-you-its-robot-even-if-it-sounds-human/ (accessed 23 December 2019).

Eisenmann C and Lynch M (2021) Introduction to Harold Garfinkel's ethnomethodological "misreading" of Aron Gurwitsch on the phenomenal field. *Human Studies* 44(1): 1–17.

Else H (2023) Abstracts written by ChatGPT fool scientists. *Nature* 613(7944): 423–423.

Ferrara E, Varol O, Davis C, et al. (2016) The rise of social bots. *Communications of the ACM* 59(7): 96–104.

Flinkfeldt M, Parslow S and Stokoe E (2021) How categorization impacts the design of requests: asking for email addresses in call-centre interactions. *Language in Society* 51(4): 693–716.

French RM (2000) The Turing Test: the first 50 years. *Trends in Cognitive Sciences* 4(3): 115–122.

Garfinkel H (1967) *Studies in Ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.

Garfinkel H (2021) Ethnomethodological misreading of Aron Gurwitsch on the phenomenal field. *Human Studies* 44(1): 19–42.

Garun N (2019) One year later, restaurants are still confused by Google Duplex. *The Verge*, 9 May. Available at: https://www.theverge.com/2019/5/9/18538194/google-duplex-ai-restaurants-experiences-review-robocalls (accessed 9 May 2019).

Goodwin C (1986) Between and within: alternative sequential treatments of continuers and assessments. *Human Studies* 9(2): 205–217.

Goodwin C (2000) Action and embodiment within situated human interaction. *Journal of Pragmatics* 32(10): 1489–1522.

Goodwin C and Heritage J (1990) Conversation analysis. *Annual Review of Anthropology* 19(1): 283–307.

Google Developers (2018) Keynote (Google I/O '18). Available at: https://www.youtube.com/watch?v=ogfYd705cRs (accessed 2018).

Harnad S (1992) The Turing Test is not a trick: Turing indistinguishability is a scientific criterion. *SIGART Bulletin* 3(4): 9–10.

Healey PGT, de Ruiter JP and Mills GJ (2018) Editors' introduction: miscommunication. *Topics in Cognitive Science* 10(2): 264–278.

Hepburn A and Bolden GB (2017) *Transcribing for Social Research*. London: Sage.

Heritage J (1997) Conversation analysis and institutional talk: analyzing data. In: Silverman D (ed.) *Qualitative Analysis: Issues of Theory and Method*. London: Sage, pp. 161–182.

Heritage J (2008) Conversation analysis as social theory. In: Bryan Turner (ed.) *The Newblackwell Companion to Social Theory*. London: Blackwell, pp. 300–320.

Heritage J and Clayman S (2010) *Talk in Action. Interactions, Identities, and Institutions*. Chichester, UK: Wiley.

Hester S and Francis D (2001) Is institutional talk a phenomenon? Reflections on ethnomethodology and applied conversation analysis. In: McHoul A and Rapley M (eds) *How to Analyze Talk in Institutional Settings: A Casebook of Methods*. London: Bloomsbury Publishing, pp. 206–217.

Hoey EM and Raymond CW (2022) Managing conversation analysis data. In: Berez-Kroeker A, McDonnell B and Koller E (eds) *The Open Handbook of Linguistic Data Management*. Cambridge, MA: MIT Press, pp. 257–266.

Hoey EM and Stokoe E (2018) Eligibility and bad news delivery: how call-takers reject applicants to university. *Linguistics and Education* 46: 91–101.

Housley W, Albert S and Stokoe E (2019) Natural action processing. In: *Proceedings of the halfway to the future symposium* (eds Fischer JE, Martindale S, Porcheron M, et al.), Nottingham, 19–20 November, 1–4. New York: Association for Computing Machinery.

Housley W, Webb H, Edwards A, et al. (2017) Membership categorisation and antagonistic Twitter formulations. *Discourse & Communication* 11(6): 567–590.

International Society for Conversation Analysis (ISCA) (2023) Schegloff media archive – ISCA. Available at: https://www.conversationanalysis.org/schegloff-media-archive/ (accessed 2 August 2024).

Ivarsson J and Lindwall O (2023) Suspicious minds: the problem of trust and conversational agents. *Computer Supported Cooperative Work (CSCW)* 32(3): 545–571.

Jefferson G (1974) Error correction as an interactional resource. *Language in Society* 3(2): 181–199.

Jefferson G (1987) On exposed and embedded correction in conversation. In: Button G and Lee JRE (eds) *Talk and Social Organization*. Clevedon: Multilingual Matters, pp. 86–100.

Joubin AA (2024) Performativity and Trans Literature. In: Vakoch DA and Sharp S (eds) *The Routledge Handbook of Trans Literature*. New York: Routledge, pp. 29–39.

Kaimaki M (2011) Transition relevance and the phonetic design of English call openings. *Journal of Pragmatics* 43(8): 2130–2147.

Kendrick KH (2017) Using conversation analysis in the lab. *Research on Language and Social Interaction* 50(1): 1–11.

Kurzweil R and Kapor M (2009) A wager on the Turing Test. In: Epstein R, Roberts G and Beber G (eds) *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*. Dordrecht: Springer, pp. 463–477.

Lee S-H (2006) Second summonings in Korean telephone conversation openings. *Language in Society* 35(2): 261–283.

Lee S-H (2011) Responding at a higher level: activity progressivity in calls for service. *Journal of Pragmatics* 43(3): 904–917.

Leviathan Y and Matias Y (2018) Google Duplex: an AI system for accomplishing real-world tasks over the phone. *Google AI Blog*. Available at: http://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html (accessed 22 May 2019).

Levinson SC (1979) Activity types and language. *Linguistics* 17(1979): 365–399.

Levinson SC (2013) Action formation and ascription. In: Sidnell J and Stivers T (eds) *The Handbook of Conversation Analysis*. Chichester: Wiley-Blackwell, pp. 101–130.

Liang W, Yuksekgonul M, Mao Y, et al. (2023) GPT detectors are biased against non-native English writers. *Patterns* 4(7): 100779.

Liesenfeld A and Dingemanse M (2024) Interactive probes: towards action-level evaluation for dialogue systems. *Discourse & Communication* 18(6): 954–964.

Loebner H (2009) How to hold a Turing Test contest. In: Epstein R, Roberts G and Beber G (eds) *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*. Dordrecht: Springer, pp. 173–179.

Lopez A, Liesenfeld A and Dingemanse M (2022) Evaluation of automatic speech recognition for conversational speech in Dutch, English and German: what goes missing? In: *Proceedings of the 18th conference on natural language processing (KONVENS 2022)* (eds Schaefer R, Bai X, Stede M, et al.), Silchar, India, 12–15 September, pp. 135–143. Vienna: KONVENS 2022 Organizers.

Maynard DW and Turowetz J (2022) *Autistic Intelligence: Interaction, Individuality, and the Challenges of Diagnosis*. Chicago, IL: University of Chicago Press.

Mlynář J, de Rijk L, Liesenfeld A, et al. (2024) AI in situated action: a scoping review of ethnomethodological and conversation analytic studies. *AI & SOCIETY* 40: 1497–1527.

Natale S (2021) *Deceitful Media: Artificial Intelligence and Social Life after the Turing Test*. Oxford: Oxford University Press.

Natale S (2023) AI, human-machine communication and deception. In: Guzman AL, McEwen R and Jones S (eds) *The Sage Handbook of Human-machine Communication*. Thousand Oaks, CA: Sage, pp. 401–408.

Natale S and Depounti I (2024) Artificial sociality. *Human-Machine Communication* 7(1): 83–98.

Oberhaus D (2018) The Story of Lenny, the Internet's Favorite Telemarketing Troll. *Vice*, 21 November. Available at: https://www.vice.com/en/article/the-story-of-lenny-the-internets-favorite-telemarketing-troll/

O'Leary DE (2019) Google's Duplex: pretending to be human. *Intelligent Systems in Accounting, Finance and Management* 26(1): 46–53.

Pearl C (2016) *Designing Voice User Interfaces: Principles of Conversational Experiences*. Sebastopol, CA: O'Reilly.

Pino M and Edmonds DM (2024) Misgendering, cisgenderism and the reproduction of the gender order in social interaction. *Sociology* 58(6): 1243–1262.

Porcheron M, Fischer JE, Reeves S, et al. (2018) Voice interfaces in everyday life. In: *Proceedings of the 2018 ACM conference on human factors in computing systems (CHI'18)*, Montreal QC Canada, 21–26 April, pp. 1–12. New York: Association for Computing Machinery.

Potter J and Hepburn A (2012) Eight challenges for interview researchers. In: Gubrium JF, Holstein JA, Marvasti AB, et al. (eds) *The Sage Handbook of Interview Research: The Complexity of the Craft*. Thousand Oaks, CA: Sage, pp. 555–571.

Psathas G (1999) Studying the organization in action: membership categorization and interaction analysis. *Human Studies* 22(2/4): 139–162.

Purver M, Hough J and Howes C (2018) Computational models of miscommunication phenomena. *Topics in Cognitive Science* 10(2): 425–451.

Pütz O and Esposito E (2024) Performance without understanding: how ChatGPT relies on humans to repair conversational trouble. *Discourse & Communication* 18(6): 859–868.

Relieu M (2024) How Lenny the bot convinces you that he is a person: storytelling, affiliations, and alignments in multi-unit turns. *Discourse & Communication* 18(6): 1–10.

Relieu M, Sahin M and Francillon A (2020) A configurational approach to conversational lures. *Reseaux* 220–221(2): 81–111.

Ruck DJ, Rice NM, Borycz J, et al. (2019) Internet research agency Twitter activity predicted 2016 U.S. election polls. *First Monday* 24(7). DOI: 10.5210/fm.v24i7.10107.

Sacks H (1984) Notes on methodology. In: Atkinson JM and Heritage J (eds) *Structures of Social Action: Studies in Conversation Analysis*. London: Cambridge University Press, pp. 21–27.

Sacks H (1995) *Lectures on Conversation* (ed Jefferson G). London: Wiley-Blackwell.

Sacks H, Schegloff EA and Jefferson G (1974) A simplest systematics for the organization of turn-taking for conversation. *Language* 50(4): 696–735.

Sahin M, Relieu M and Francillon A (2017) Using chatbots against voice spam: analyzing Lenny's effectiveness. In: *Proceedings of the thirteenth USENIX conference on usable privacy and security* (eds Zurko ME, Chiasson S and Smith M), Santa Clara CA, pp. 319–337. Berkeley, CA: USENIX Association.

Schegloff EA (1968) Sequencing in conversational openings. *American Anthropologist* 70(6): 1075–1095.

Schegloff EA (1986) The routine as achievement. *Human Studies* 9(2): 111–151.

Schegloff EA (1987) Analyzing single episodes of interaction: an exercise in conversation analysis. *Social Psychology Quarterly* 50(2): 101–114.

Schegloff EA (1992) Repair after next turn: the last structurally provided defense of intersubjectivity in conversation. *American Journal of Sociology* 97(5): 1295–1345.

Schegloff EA (1993) Reflections on quantification in the study of conversation. *Research on Language & Social Interaction* 26(1): 99–128.

Schegloff EA (2009) One perspective on conversation analysis: comparative perspectives. In: Sidnell J (ed.) *Conversation Analysis*. Cambridge: Cambridge University Press, pp. 357–406.

Schegloff EA (2010) Some other "uh(m)s." *Discourse Processes* 47(2): 130–174.

Schegloff EA (2015) Conversational interaction: the embodiment of human sociality. In: Tannen D, Hamilton HE and Schiffrin D (eds) *The Handbook of Discourse Analysis*. 2nd ed. Hoboken, NJ: John Wiley, pp. 346–366.

Schegloff EA, Jefferson G and Sacks H (1977) The preference for self-correction in the organization of repair in conversation. *Language* 53(2): 361–382.

Schweizer P (1998) The truly total Turing Test. *Minds and Machines* 8(2): 263–272.

Scott R (Director), Fancher H (Writer), Peoples D, et al. (1982) *Blade Runner* (Motion picture). West Hollywood, CA: *The Ladd Company*.

Shen Z, Wang K, Zhang Y, et al. (2024) Combating phone scams with LLM-based detection: where do we stand? *arXiv:2409.11643. arXiv*. Available at: http://arxiv.org/abs/2409.11643

Sidnell J (2014) The architecture of intersubjectivity revisited. In: Enfield NJ, Kockelman P and Sidnell J (eds) *The Cambridge Handbook of Linguistic Anthropology*. Cambridge: Cambridge University Press, pp. 364–399.

Steensig J and Heinemann T (2015) Opening up codings? *Research on Language and Social Interaction* 48(1): 20–25.

Stivers T (2015) Coding social interaction: a heretical approach in conversation analysis? *Research on Language and Social Interaction* 48(1): 1–19.

Stokoe E (2013) The (in)authenticity of simulated talk: comparing role-played and actual interaction and the implications for communication training. *Research on Language & Social Interaction* 46(2): 165–185.

Stokoe E, Albert S, Buschmeier H, et al. (2024) Conversation analysis and conversational technologies: finding the common ground between academia and industry. *Discourse and Communication* 18(6): 837–847.

Stokoe E, Albert S, Parslow S, et al. (2021) Conversation analysis and conversation design: where the moonshots are. *Medium*, 3 June. Available at: https://elizabeth-stokoe.medium.com/conversation-design-and-conversation-analysis-c2a2836cb042

Stokoe E and Richardson E (2023) Asking for help without asking for help: how victims request and police offer assistance in cases of domestic violence when perpetrators are potentially co-present. *Discourse Studies* 25(3): 383–408.

Stokoe E, Sikveland RO, Albert S, et al. (2020) Can humans simulate talking like other humans? Comparing simulated clients to real customers in service inquiries. *Discourse Studies* 22(1): 87–109.

Stokoe E, Sikveland RO and Symonds J (2016) Calling the GP surgery: patient burden, patient satisfaction, and implications for training. *British Journal of General Practice* 66(652): e779–e785.

Suchman L (2023) The uncontroversial "thingness" of AI. *Big Data & Society* 10(2). DOI: 20539517231206794.

Turing A (1950) Computing machinery and intelligence. *Mind* 49: 433–460.

Turowetz J and Rawls AW (2021) The development of Garfinkel's 'Trust' argument from 1947 to 1967: demonstrating how inequality disrupts sense and self-making. *Journal of Classical Sociology* 21(1): 3–37.

Wallace RS (2009) The anatomy of A.L.I.C.E. In: Epstein R, Roberts G and Beber G (eds) *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*. New York: Springer, pp. 181–210.

Whalen J, Whalen M and Henderson K (2002) Improvisational choreography in teleservice work. *The British Journal of Sociology* 53(2): 239–258.

Whalen MR and Zimmerman DH (1987) Sequential and institutional contexts in calls for help. *Social Psychology Quarterly* 50(2): 172.

Whitby B and Oliver K (2000) How to avoid a robot takeover: political and ethical choices in the design and introduction of intelligent artifacts. *Quarterly Journal of the Society for the Study of Artificial Intelligence and the Simulation of Behaviour*, 104.

Wilkinson R, Rae J and Rasmussen G (eds) (2020) *Atypical Interaction: The Impact of Communicative Impairments within Everyday Talk*. London: Palgrave Macmillan.

Wilkinson S and Kitzinger C (2006) Surprise as an interactional achievement: reaction tokens in conversation. *Social Psychology Quarterly* 69(2): 150–182.

Zhan X, Xu Y and Sarkadi S (2023) Deceptive AI ecosystems: the case of ChatGPT. In: *Proceedings of the 5th International Conference on Conversational User Interfaces*, Eindhoven, 19–21 July, pp. 1–6. New York: Association for Computing Machinery.

Author biographies

Saul Albert is a Senior Lecturer in Social Science (Social Psychology) in the Department of Communication and Media at Loughborough University, UK, and a member of the Discourse And Rhetoric Group. His research at the intersection of cognitive science and conversation analysis explores how efforts to manage miscommunication underpin human social interaction.

William Housley is Professor of Sociology at Cardiff University, School of Social Sciences, Wales, UK. He was lead editor of the recently published *SAGE Handbook of Digital Society* (Sage 2022) and has published numerous papers and books including *Interaction in Multidisciplinary Teams* (Routledge 2017) and *Society in the Digital Age: An Interactionist Perspective* (Sage 2021).

Rein Ove Sikveland is Professor at the Centre for Academic and Professional Communication (SEKOM) at the Norwegian University of Science and Technology (NTNU), Trondheim. Rein is a conversation analyst and phonetician. He has published several papers on the use of phonetic detail and prosody in conversations, and has applied conversation analytic research to education, health services and crisis negotiations.

Elizabeth Stokoe is a Professor in the Department of Psychological and Behavioural Science at The London School of Economics and Political Science, UK. Her research in conversation analysis and membership categorization currently focuses on the ways that 'conversation' is leveraged in both research and industry processes and products.