

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/124584/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Stokoe, Elizabeth, Sikveland, Rein, Albert, Saul, Hamann, Magnus and Housley, William 2020. Can humans simulate talking like other humans? Comparing simulated clients to real customers in service inquiries. *Discourse Studies* 22 (1) , pp. 87-109. 10.1177/1461445619887537

Publishers page: <http://dx.doi.org/10.1177/1461445619887537>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



Can humans simulate talking like other humans?

Comparing simulated clients to real customers in service inquiries

Elizabeth Stokoe¹, Rein Sikveland², Saul Albert³, Magnus Hamann⁴ *and* William Housley⁵

Abstract

How authentic are inquiry calls made by simulated clients, or ‘mystery shoppers’, to service organizations, when compared to real callers? We analysed 48 simulated and 63 real inquiry calls to different veterinary practices in the UK and Ireland. The data were transcribed for conversation analysis, as well as coded for a variety of call categories including reason for the call, call outcome, and turn design features. Analysis revealed systematic differences between real and simulated calls in terms of 1) reasons for the call, call outcome, and call duration, and 2) how callers refer to their pets in service requests and follow-up questions about their animal. Our qualitative analyses were supported with statistical summaries and tests. The findings reveal the limitations of mystery shopper methodology for the assessment of service provision. We also discuss the implications of the findings for the use of simulated encounters and the development of conversational agents.

Keywords: Simulated clients, mystery shoppers, conversation analysis, veterinarian practice, conversational agents, service encounters, requests

Introduction

The objective of this paper is to investigate how the communicative practices of simulated clients, or ‘mystery shoppers’, making telephone calls to service providers, compare to actual service users. Mystery shopping is “a form of participant observation”, in which mystery shoppers “act as customers or potential customers to monitor the quality of processes and procedures used in the delivery of a service” (Wilson, 1998, p 414). Being a simulated client involves ‘passing’ (Garfinkel, 1967): the practices through which people produce authentic

¹ ^{1,2,3,4} School of Social Sciences, Brockington Building, Loughborough University, Loughborough, LE11 3TU. Tel: 01509 263171; ⁵ School of Social Sciences, Glamorgan Building, Cardiff University, King Edward VII, Cardiff, CF10 3WT. Tel: 029 2087 5179

and unaccountable social activity. As this paper will demonstrate, passing as an authentic pet owner during service calls to a veterinary practice involves a practical and granular display of stake and interest, which mystery shoppers do not accomplish, at least in our data.

The paper is situated in multiple research contexts. First, there is much commentary and analysis of the wider purpose and ethical practice of mystery shopping. Second, researchers have examined the evaluations made by mystery shoppers of the services they engage with, including the validity of such a customer service methodology. Third, there is a related literature on simulated role-play, in which actors (or other professionals) play the part of customers, patients, suspects, and so on, as part of communication skills training and assessment. Finally, within the simulated client/mystery shopping literature, a handful of studies tests the authenticity or otherwise of the actions and practices done by the simulated client when compared to real service users.

In this paper, we make a practical distinction between two types of simulated client: mystery shoppers and role-play. In mystery shopping, only one interlocutor knows that the encounter is ‘simulated’. In role-play, both parties know that the situation is simulated. While scant research examines actual interaction in simulation or assesses its authenticity, this paper is, to the best of our knowledge, the first to scrutinize the unfolding empirical reality of mystery shopping. We also make a direct comparison to real users of the same service, to address our core research question regarding a mystery shopper’s ability to generate authentic action (e.g., requests for service).

MYSTERY SHOPPING AND CUSTOMER SATISFACTION

A great deal is written about the history, practice, and use of mystery shopping in the management and business literatures. Mystery shopping is a large industry, itself provided as a service and used by organizations to evaluate service quality. It is founded on a range of assumptions, including the importance of customer satisfaction to the performance of an organization. Furthermore, the perceived value of mystery shopping as a way of testing service provision assumes that customer experience is as important to customer satisfaction as the product or tangible benefit provided (Bateson, 1992). As Wilson (1998) notes, mystery shopping is designed to evaluate the *process* of service. This stands in contrast with post-hoc customer surveys and its ubiquitous contemporary methodology, the single-item ‘Net Promotor Score’ (NPS: for an overview, see Owen, 2019). Addressing marketing executives and managers, Fisher and Kordupleski (2018) are critical of NPS as a proxy of customer satisfaction, for the simple reason that while it gives “an indicator of how you are doing”, it “provides no data to help you know what to do” (p. 139). This is the problem that mystery shopping seeks to address:

The emphasis is on the service experience *as it unfolds*, looking at which activities and procedures do or do not happen rather than gathering opinions about the service experience ... customer satisfaction surveys on their own [do] not provide sufficiently detailed information to allow management to identify and correct weaknesses in the service delivery process: Customers *don’t remember the detail of the service encounter*,

they can only give *an overall impression* of the service (Wilson, 1998; emphasis added).

Mystery shoppers aim to collect concrete data about different aspects of a live, unfolding customer experience, such as how long it takes a company to answer the phone, how long a queue is, or how service providers greet customers (Wilson, 1988). The methodology has proliferated in recent years, to provide not just market research for private sector commercial organizations but also public sector health care, where it is used extensively (Jacobs et al, 2018). In a study of the provision of emergency contraception in Scotland, for example, Glasier et al (2010) showed that few pharmacists followed Government guidance to advise patients about future contraceptive use. Similarly, in a study of the selling of ‘restricted medicines’ in New Zealand, Norris (2002) observed that aspects of guidance were not followed. She found that mystery shoppers struggled to distinguish between pharmacists (who were qualified to sell restricted drugs) and other staff (who were not), and that the former were only involved in transactions in 46% of visits.

Problems with mystery shopping methodology

In a review of over 60 mystery shopper studies, Jacobs et al (2018) made a series of criticisms of the methodology. First, they argued that sample size and number of visits were often too small to generalize across an organization’s practice, raising questions about the validity of any observations undertaken (Calvert, 2005). They also argued that, like post-hoc reporting on a customer survey, mystery shoppers rely on their memory, which may be biased by variables like gender, age, dress, or other factors, further reducing the validity of their evaluation. Third, Jacobs et al raise the issue of the ethics of mystery shopping methods, given that informed consent is generally absent. We return to this point later when describing the dataset used in the current study.

A further problem with the pharmacy examples described above is the assumption that mystery shoppers produce requests (and other actions) in the same way that patients who actually want drugs (and other services) do. In his interviews with managers, Wilson (1998) reported that mystery shoppers “must match a customer profile that is appropriate for the scenario that they are being asked to enact”. If mystery shoppers are unable to reproduce a genuine service encounter, then their reports back to the organization are, at best, reduced in validity and, at worst, potentially damaging for the employees who have interacted with them. In their review, Jacobs et al (2018) also highlight potential authenticity issues:

[The] mystery customer as such – their knowledge and past experiences – can *influence how the interaction will unfold*. Someone who has already worked in the evaluated sector may see things differently from a person with no prior knowledge or experience in the sector (Calvert, 2005). ... there may be a discrepancy between the pre-set scenario – a prerequisite of mystery shopping – and *its actual application in the course of the study* (emphasis added).

We know little about ‘how the interaction will unfold’ and the ‘actual application’ of mystery shopping. The value of analysing real service users is that it lets us examine the core

phenomena that mystery shoppers attempt to elicit and analyse: the unfolding customer experience. However, our analysis reveals how the lack of authenticity in interaction places key limitations on the mystery shopping methodology. This paper focuses on the ecological validity of mystery shopping methods in particular, and in this way contributes to studies of the authenticity of simulated encounters in general.

The problem of authenticity in simulated encounters

Much is written about simulated clients and role-play methodology, and this research is used to train and assess communication skills and related competences in institutional settings of all kinds. Simulation methods involve people-in-training, from call-centre workers and corporate business managers to doctors and police officers, interacting with actors or other simulated interlocutors, using “narrative adaptations” of hypothetical or actual scenarios as the basis for the encounter (Van Hasselt et al, 2008, p.254). In the commercial world, its authenticity is treated in the literature as a non-issue. Role-play is simply assumed to be authentic, or ‘authentic enough’, to train people to have better interactions and assess how they communicate, extrapolating from the training room to the workplace (Stokoe, 2013). Research about authenticity largely addresses questions about which category of ‘role-player’ provides for the most realistic experience (comparing, say, actors with other trainees or, in medical training, ‘standardized patients’, e.g., Mounsey et al, 2006), and those which ask participants to report, post-hoc, on the perceived authenticity of training encounters (e.g., Bokken et al, 2009).

Stokoe (2013) describes problems in making simulated encounters more authentic, and the more general presumption that authentic turns of talk can be invented based on vernacular understandings of interaction. Indeed, conversation analysts have shown how studies of talk yield “results at variance with our common-sense intuitions about how some action is accomplished” (Schegloff, 1996, p. 166-169). Furthermore, what is at stake in simulation is different from what is at stake in real encounters. For those having their communication skills evaluated, it is their performance and ‘score’ as trainees that are at stake rather than, for example, the healthcare outcomes in a real consultation (De la Croix & Skelton, 2009; see also Pilnick et al, 2018). Atkins et al (2016) discuss whether simulated encounters can and should function as a proxy for the real given their potential to disadvantage candidates trained overseas. They write:

Candidates who can handle the social and linguistic complexities of the artificial context of assessed role-plays score highly – yet what is being assessed is not real professional communication, but the ability to voice a credible appearance of such communication... Fidelity may not be the primary objective of simulation for medical training, where it enables the practising of skills. However, the linguistic problems and differences that arise from interacting in artificial settings are of considerable importance in assessment, where we must be sure that the exam construct adequately embodies the skills expected for real-life practice.

Some of the challenges of using simulated clients – whether participants know they are doing role-play or are interacting with a mystery shopper – are the same. The person whose skills

are being assessed may be interacting with someone whose own ability to produce an authentic performance is untested. In her analysis comparing police officers interviewing real suspects to their interactions in training with actors, Stokoe (2013) showed that officers formulated core actions (e.g., introducing themselves; reminding suspects of their legal rights) differently. Drawing on the same dataset, Stokoe and Sikveland (2017) further showed that real suspects and actors respond differently to questions from police officers, whose questions were also often formulated differently for each group. Atkins (2018) has reported similar and consequential differences in

The aim of this paper is to further explore how human interlocutors accomplish – or fail to accomplish – the job of being an authentic party to an interaction. That is, when parties have a particular stake in the actions they initiate and pursue, does their stake in the encounter matter for the design, sequential location, and trajectory of those actions? Following the analysis, we will discuss the implications for mystery shopping as a market research methodology, as well as broader ramifications for understanding and using simulated conversations for a variety of purposes.

DATA AND METHOD

We analyse two datasets: 48 telephone calls from simulated clients, or mystery shoppers, to different veterinarian practices in the UK and Ireland in 2016, and 63 calls from real pet owners. The mystery shopper calls ranged from 20 seconds to two minutes and 36 seconds; real calls from 20 seconds to six minutes and 29 seconds. The calls were provided by a UK-based specialist customer experience organization with a view to identifying the communicative practices that are effective to underpin training for vet receptionists. Participating organizations consented to provide their calls that, like most service organizations, are already recorded for training and evaluation purposes. Data were supplied on the basis that recordings were stored securely using encryption and, when used for research and training, fully anonymized in terms of both audio file (transforming voice pitch and removing identifying service, pet and owner names, places, and so on) and transcript (using pseudonyms).

Within the marketing and business literature, several authors have raised ethical concerns about the mystery shopper methodology. These concerns include the absence of informed consent – not about recording of calls, necessarily, but that service providers are being in some way deceived by the ‘customer’. Relatedly, one may regard mystery shopping as an ‘exploitative fabrication’ (Goffman, 1974, p. 103), with potentially negative consequences for the employees or managers of organizations who are ‘duped’ by mystery shoppers. Some argue that “informing the evaluated staff is a prerequisite for the conduct of the study, without obviously specifying the exact moment of the interactions” (Jacobs et al, 2018). Importantly in our study, we are not reporting back about the ‘performance’ of any party to the mystery shopping interaction; neither the service provider nor the mystery shopper themselves. Rather, the remit of this research, initiated by the customer experience organization, was to examine real calls and use those to develop evidence-based training instead.

All data were transcribed verbatim in order to identify core features and establish trends in the datasets. We transcribed all extracts for conversation analysis (Hepburn & Bolden, 2017). Conversation analysis focuses on the systematic identification of composite actions in encounters, such as advising, informing or questioning, and their design and organizational structure. It is guided by an incremental process of phenomenon and hypothesis-formation, rather than a hypothetico-deductive research design (De Ruiter & Albert, 2017). Because this is a comparative study, we supplemented our qualitative analysis with a coding scheme (Stivers, 2015) and used chi-squared tests to check for the statistical independence of the categories. We included these tests to strengthen claims about distributional relationships between the datasets. We coded for reason for the call, call outcome, call length, as well as for potentially systematic differences in turn design and action formation.

ANALYSIS

The first section of analysis examines differences in the reason for the call, call outcome, and call duration; the second reports on fine-grained turn design and sequence organizational features, including the way requests are built by real customers (RC) and mystery shoppers (MS), including how pets are referred to, and differences in the way pet information (e.g., weight, age, name, breed) becomes relevant to the unfolding interaction, either offered by the caller or in response to vet receptionists' questions.

1. Key differences between real customer and mystery shopper calls

a. Reason for the call

A fundamental difference between mystery shopper (MS) and real caller (RC) calls is that they involve different kinds of activities. In institutional and ordinary talk, it is a robust finding of conversation analysis that callers articulate a 'reason-for-the-call' (Sacks, 1992) in or near their first speaking turn, following greetings and identification. These core components of social interaction, and their presence and absence, establish 'the routine as achievement' (Schegloff, 1986). Parties orient the presence/absence of greetings ("Hello? ... Hello...?"), reasons for the call (e.g., "Anyway, I just called to..."; "the reason I'm calling is"), and so on. In our data, both RC and MS callers performed the reason-for-the-call *action* at the start of calls, in its apposite slot. However, reasons differed. As Couper-Kuhlen (2001) explains, the 'reason-for-the-call' "not only provides an account for a particular occasion of talk, it also has consequences for subsequent talk, in that the identities and relevancies it establishes condition appropriate ways of responding" (p. 4).

Extract 1 is an example of a real customer (RC) request in the 'reason for the call' slot:

Extract 1: RC-Other 12*

01 V: Good evening Vet Centre,=Jackie speaking,=how can I help.
 02 C: Hello?=>I wonder if it's possible to make an appointment<
 03 <for my cat> tomorrow=for a >follow up< uhm he's had an
 04 operation?
 05 V: Okay? What was the surname sorry?

The caller requests an appointment, and accounts for their request, citing pet illness (lines 03-04). The receptionist begins to progress this request, by first soliciting the caller's name. In contrast, in Extract 2, a mystery shopper (MS) caller makes a request for information.

Extract 2: MS-Jabs 2
 01 V: Hello,=Vet Clara Dee, Can I help you,
 02 (0.4)
 03 C: .thhh hiya, Yeah.=I just ↑wanted to know how much you charge
 04 for uh puppy: vaccinations.
 05 (0.5)
 06 V: Uh fifty five euros covers the two vaccinations.

Analysis of the datasets revealed that RC callers mostly called to make appointments for their pets while MS callers asked, without this being scriptedⁱ, almost exclusively for information about a service price (e.g., vaccinations) or process (e.g., euthanasia). These differences were consequential for the trajectory of the call, as they set up opportunities (or 'response slots') for the interlocutor to take different next actions (see Stivers, 2012). Note that, in Extract 2, the MS caller does not include an account for calling, unlike the RC caller in Extract 1. Figure 1, below, however, summarizes this first gross difference in 'reason for the call' between the two datasets.

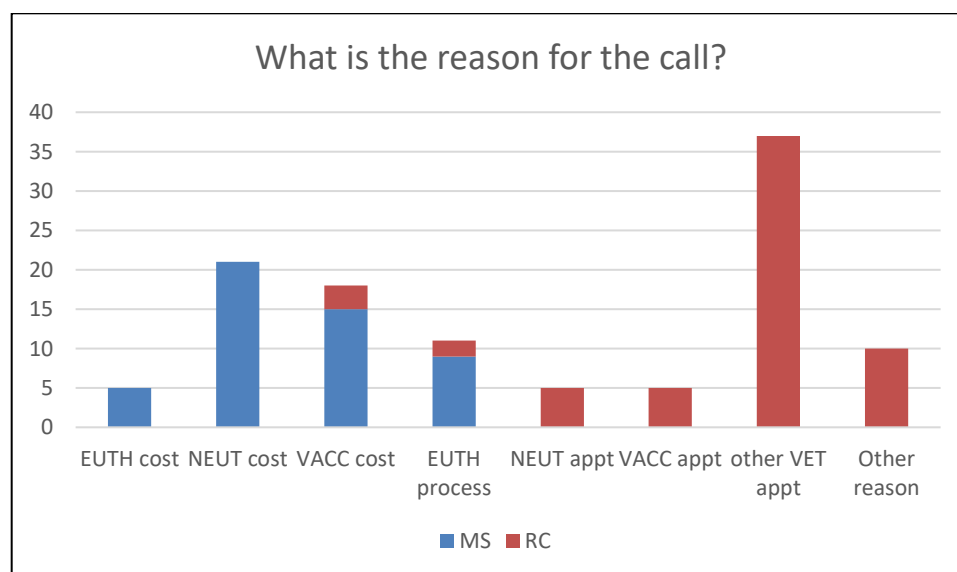


Figure 1. Distribution (n) of mystery shopper (MS) and real call (RC) call types: calls about 'cost', appointment ('appt') and 'process'. EUTH = Euthanasia; NEUT = Neutering; VACC = Vaccinations

Figure 1 shows that, while MS callers vary the type of service they are enquiring about (neutering, vaccinations, euthanasia), the overall request is for information about the *cost* of those services.

Extract 1 is typical of another observation that RC callers often formulate an account for calling which includes some detail about their pet's illness. For instance, in Extract 3, the caller requests an appointment, and couches her request in a turn that also comprises a concern about her puppy.

Extract 3: RC_Other 11

01 C: Hi Jackie=I was wondering- I'm a bit worried about my puppy
02 (an- our) ↓dog and I'm just wondering if I could bring her in
03 for an emergency appointment.
04 V: °Okay=just bear with #me a minute#°

Note that the caller expresses 'worry' as part of her reason for the call. As well as finding that MS callers never make appointments, neither do they include information about the animal's state of health in their reason for the call. We do not have any comparable calls in the MS dataset in which expresses 'worry' about their animal. However, consider Extracts 4 and 5, which contrast RC and MS calls in which euthanasia is the reason for the call. Extract 4 is a RC call; the recording begins just after the receptionist has answered the phone.

Extract 4: RC-End of life 2

01 C: ~.hh hh~
02 (0.7)
03 V: Hello?
04 C: .shih.
05 (0.4)
06 C: H- uHello: ↑is uhm< Sandra #there# please,
07 (0.6)
08 V: Uhm She's consulting at the moment,
09 C: .snh. ~She's just called me to say that she's gonna put (.)
10 uh- that she needs to put my uhm rabbit t- (.) so I'm
11 wondering if I could be ~↑there?~

Note that the caller displays audible sounds of upset and difficulty, including those identified by Hepburn (2004) such as wobbly voice, croaky voice, sniffs, aspiration, and high pitch, in various combinations throughout her turns. The sequential placement of "H- uHello:", several lines into the interaction (line 06) and not in the usual greetings-identification slot, combined with these sounds, mark out the caller as in distress. Note also that, when it comes to formulating the reason for the call (lines 09-11), the caller's turn contains numerous perturbations including repair initiators and hesitations, and she does not articulate the word 'sleep'. Instead, she cuts off on 'to', and does not complete this idiomatic way of referring to euthanasia.

Extract 5 is from the MS dataset, and the caller is also inquiring about putting an animal 'to sleep'.

Extract 5: MS-Euth 2

01 V: Good afternoon.=Loddington Pets,=Jan speakin:, how c'n help.
02 (0.7)
03 C: Hello:: um:: yeah I wonder if you could:, .hh >I w's won'ring
04 (0.4) um I s- uh hh .hh I think I might need t'have my dog put
05 to sleep soon? an'- an' I was wondering how much it- it would
06 cost.

The MS caller delivers his request with some features that indicate its delicacy or trouble (hesitations, repair initiators), but no wobbly voice, sniffs, croak, or high pitch. The turn-

initial greeting is in its regular position at the start of the caller's first turn, and the caller fully articulates the idiomatic expression "put to sleep" (lines 04-05), that the RC caller did not produce. And, as noted earlier, the reason for the call is a "how much it- it would cost" inquiry (lines 05-06).

Although we do not have sufficient euthanasia calls in our dataset for thorough comparison, our analysis suggests that MS callers do not provide receptionists with the same tasks as RC callers, nor in the same terms. Mystery shopping is intended to mimic RC calls, but these fundamental differences in 'reason for the call' are felt across the dataset in myriad ways. The final example in this section highlights our observation that differing reasons for the call produce differing interactional landscapes, with different 'slots' for each party to act. For instance, since MS callers do not request appointments, they do not put receptionists in the position of, for example, responding to a request to cancel, as in Extract 6.

Extract 6: RC Vaccine 32

01 V: Good afterno:on, Johnson Veterinary Centre, Joan
 02 speaking?=↑how c'n I ↓help.
 03 C: .pt Oh hello Joan, uhm >I've got an appointment< booked
 04 for <Spotty Jackson to see:> for a nurse clinic (.) at
 05 six tonight (.) it's f'r socialisation, .hhh
 06 V: Okay?
 07 C: I have had to make an emergency appointment for the
 08 doctor's for me: =and it['s the only time I could ge:t.
 09 V: [Aw:
 10 C: .mt I have got an infected (.) insect bite from d-
 11 from walking the dog? (.) £B[elieve it or not£.
 12 V: [Oh my god (.) that's
 13 awful.

The RC caller initiates a type of request that MS callers never do, beginning with a pre-sequence in which she announces an already-made appointment (lines 03-05). Since MS callers make initial rather than follow-up calls, this kind of pre-sequence, and discussions about already-made appointments (here, "f'r socialisation,"), do not appear in our MS dataset. Furthermore, these actions lead to other actions that are then 'missing' from MS calls. For example, following a go-ahead from the receptionist ("Okay?", line 06), she provides an account for her upcoming cancellation, thus *treating* cancelling an appointment *as* accountable. Note the details of her account and its expansion across several turn constructional units (TCUs), which includes the pressing, unpredictable, immediate, and somewhat ironic nature of the situation – getting an infected insect bite while "walking the dog" ("I have had to make *an emergency appointment*"; "It's the only time I could get"; "£Believe it or not£."). Note also how the caller initiates her first turn ("Oh hello Joan") implies an ongoing relationship with the practice.

The outcome of the call in Extract 5 was to cancel the appointment. While these kinds of accounts are unremarkable, they afford opportunities to inspect interactional work only found in the RC calls dataset. For example, by providing this account for cancelling an appointment, slots are opened up in the unfolding sequence for the receptionist to respond in ways that, again, are not present in MS interactions (e.g., surprise/shock: "Oh my god";

empathy: “that’s awful”). These structural differences between MS and RC hinge on the ‘reason for the call’. In the following sections we look at how these differences lead to opportunities for other kinds of differences to emerge – both structural and in terms of call outcomes.

b. Call duration

We found a clear difference in the durational distributions of the two datasets. Mystery shopper (MS) calls were significantly shorter than real customer calls (RC). The mean duration of RC calls was 0:01:50 and 0:01:18 for MS calls². The standard deviations were 0:01:15 and 0:00:38 respectively for RC and MS calls. The wider spread of durations for RC calls is illustrated via boxplot in Figure 2.

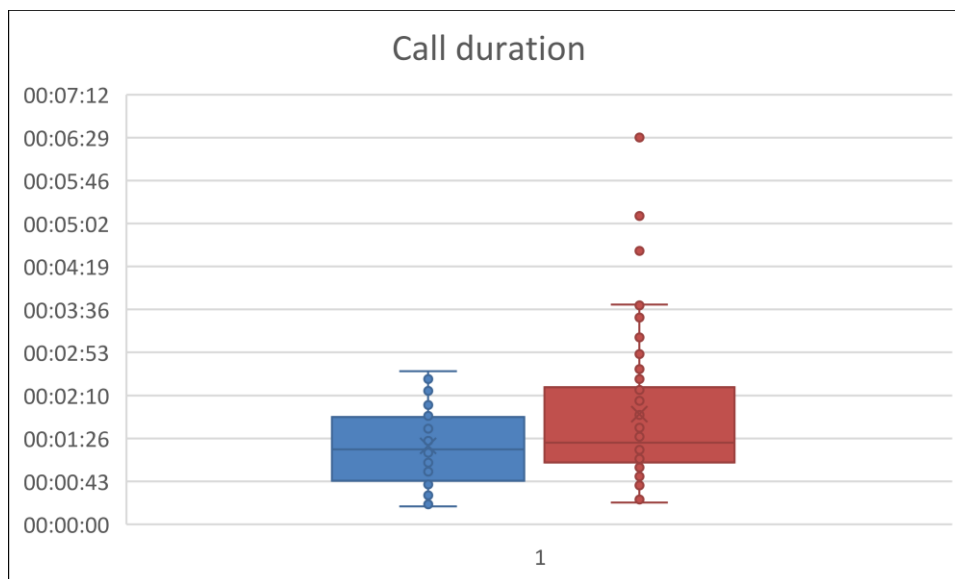


Figure 2. Boxplot of call durations in mystery shopper (blue) and real calls (red). Horizontal lines represent the respective medians; vertical lines represent range of durations.

While most RC calls, like MS calls, are below two minutes, we only find outliers of long calls up to 6.5 minutes in the RC corpus. Call length therefore provides a clear ‘tell’ that the caller is a mystery shopper. However, that calls are ‘longer’ or ‘shorter’ tells us little about their constituent activities or why they vary. Qualitative analysis reveals that, depending on the ‘reason for the call’, subsequent actions become relevant or are curtailed, and even tiny differences such as the inclusion of ‘Joan’ in Extract 6 adds to the length of real calls.

c. Call outcome

Given that real callers most commonly make appointments, an appointment is the most common outcome. By contrast, MS calls routinely end with no further action, and callers resist offers of appointments. RC callers may also resist appointment offers but are rarely *offered* them because receptionists usually progress *caller-requested* appointments. Extract 7 is an example.

² ($t = -2.891$, $df = 94.7$, $p < 0.005$)

Extract 7: RC 12 Other

01 V: Good evening Vet Centre=Jackie speaking,=how can I help.
02 C: Hello?=>I wonder if it's possible to make an appointment< for
03 my cat> tomorrow.=for a >follow up< uhm- he's had an
04 operation?
05 V: Okay? What was the surname sor[ry?
06 C: [Uhm Smithdon.
07 (1.2)
08 V: A:nd the cat's name?
09 C: Fluffy.
10 (0.5)
11 V: #Fluffy was that#. =yeah?
12 C: =Yes::h.
13 (0.6)
14 V: Let's have a lo:ok,
15 (3.8)
16 V: M- m- (0.3) ↑m::: (0.6) .mt O:kay?
17 (1.7)
18 V: That's fine=I'll just see what we've got <available>
19 tomorrow?
20 (4.6)
21 V: .hh #We can do#?
22 (1.9)
23 V: °>Just seeing if any of the° vets< #that# you've #seen before
24 are in,#
25 (1.2)
26 V: °Great (.) #Okay#° hh uh Ten fifteen Tomorrow morning?
27 would [that would be ok[a:y?
28 C: [That would- [that would be lovely yes.
29 V: Ye[ah
30 C: [Thank you.
31 V: #Lovely#.
32 V: h #↓Perfect#=That's all popped in=We'll see Fluffy at ten
33 fifte:en.
34 C: Fantastic.=Thank you very much.
35 V: No worries=you're we[lcome.
36 C: [Cheers. Thank you
37 V: [Bye
38 C: [Thanks
39 V: By:e

Having made a request for an appointment, the call proceeds with a series of fitted, apposite, and preferred turns³ – an offer and acceptance, followed by confirmation of the appointment. The call ends via stepwise series of appreciations (lines 30, 34) and assessments (lines 28, 34) as well as reciprocal closing ‘thanks’ and ‘byes’ that comprise an archetype closing (e.g., Schegloff & Sacks, 1973).

Extract 8 illustrates a typical MS call outcome.

Extract 8: MS Jobs 2

³ A ‘preferred’ turn is, in brief, one that progress the course action currently underway, see Pillet-Shore (2017) for a full explanation.

01 V: Hello,=Vet Clara Dee, Can I help you,
 02 (0.4)
 03 C: .thhh hiya, Yeah.=I just ↑wanted to know how much you
 04 charge for uh puppy: vaccinations.
 05 (0.5)
 06 V: Uh fifty five euros covers the two vaccinations.
 07 (0.6)
 08 C: Fifty five. <Okay [cool.
 09 V: [#Yeah(p) #.
 10 (0.4)
 11 V: Do you want me to book you in for a day?
 12 (0.6)
 13 C: Uhm: probably not right now but I might call back later if
 14 that's o[kay,]
 15 V: [↓O k]ay.
 15 V: All right, That's [↑fine,
 16 C: [Thank you very [much.
 17 V: [↑Thank you,
 18 V: ↑Bye bye.
 19 (0.3)
 20 C: Bye.

The receptionist provides the requested information (line 06), which the caller confirms as a sequence-closing third turn (line 08) followed by a fourth position confirmation from the receptionist (line 09). At line 10, a silence opens up. The MS caller does not do what might be expected next: make an appointment for her puppy's vaccinations. The receptionist treats this as the relevant next action by offering to make an appointment instead ("Do you want me to book you in for a day?", line 11). In response to this offer, the caller produces a turn containing the classic features of dispreference: it is delayed, begins with a hesitation, and is a weaker rejection than a flat 'no' ("probably not right now"). At line 13, the caller accounts for her imminent exit from the call, suggesting she "might call back later", effectively forestalling further offers of help. To the best of our knowledge, mystery shoppers do not subsequently call back. While the real caller in Extract 7 also uttered 'thank you very much' towards the end of the call, in that context 'thank you' functions as an appreciation – the caller's second another 'thank you' collaboratively ends the call with the receptionist. In Extract 8, 'thank you very much' occupies a closing slot only.

This section has charted differences between real customer and mystery shopper calls in terms of reasons for the call (RC callers request appointments; MS callers request information), call duration (RC calls are longer), and subsequent call outcomes. Next, we examine the detail of callers' requests and how they refer to their pets and respond to questions from receptionists.

2. Differences in request design and responses to follow-up questions

This paper asks, "Can humans simulate other humans?". So far, our analysis has shown that, when calling for service, MS callers – who may or may not have a pet, but not the pet they are calling about – initiate different types of requests, in calls that have different outcomes,

compared to RC callers. In this section, we examine some features of the turn design of calls, including pauses and hesitations, and how they to occur differently in MS and RC calls.

a. How pets are referred to in requests

We return to callers' 'reason for the call', but this time examine the way pets are referred to in request slots – a recurrent feature across the data. Consider three extracts from RC calls.

Extract 9: RC Injury 1

01 C: Could I make an appointment for my dog please? He's had (.)
02 (as) a bad ear.

Extract 10: RC Other 1

01 C: .h Hello::.=I:t's Joanna Clarke here.=I'm ringing regarding
02 um my lab Brownie, .hh he:'s (0.2) off his food,

Extract 11: RC Vaccine 5

01 C: .ptk uhm I need to make an appointment for: uh both my cats
02 to come around for their booster injections,

In Extracts 9-11, RC callers refer to the object of their requests – their pet – in ways that follow previously observed rules for referring to absent third parties. That is, on first mention, they use what conversation analysts call a 'locally initial' reference (e.g., Schegloff, 1996b) rather than, say, refer to their pets just by their names. For example, if callers said, "I'm ringing regarding Brownie", it may not be as recognizable as "my lab Brownie". The rules of recipient design reveal themselves as callers formulate their relationship with the absent third party on whose behalf they are calling (e.g., "my dog") as part of their request. This echoes other appointment-making practices in which callers ask to make an appointment "for my husband" and not "for James", using a reference which is recipient-designed to be recognized by the receptionist and avoid the need to ask, "Who's James?"

In each of Extracts 9-11, callers formulate their pet using the possessive pronoun *my*: "my dog"; "my lab"; "my cats". Furthermore, in providing an account for the requested appointment ("an operation"; "a bad ear"; "off his food"; "booster injections"), RC callers supply the receptionist with information. This information is provided within the same turn and is 'volunteered' rather than produced in response to a question from the vet. In Extract 10, the caller also supplies the breed and name of her dog ("my lab[rador] Brownie") without being asked to do so. Supplying the name of their pet was common across the RC calls dataset. In Extract 12, the caller anticipates the relevance of their pet's name.

Extract 12: RC-Other 3

01 V: =What's the surname?
02 (0.3)
03 C: Uh Jakobsen [spelling name]
04 (1.4)
05 C: [And it's for Blackie.
06 V: [And your p-

Following the receptionist's query about the caller's name (line 01), which the caller provides, the caller then supplies her pet's name ("And it's for Blackie") in anticipation of the receptionist's next action and the fitted sequential location for it (line 05). In overlap, the receptionist begins to solicit the pet's name, but she abandons her turn as the caller supplies the information (line 06). We coded the datasets according to whether the pet's name was supplied by the caller as part of their request, or 'voluntarily' in a post expansion sequence such as in Extract 12. Figure 3 shows that RC callers supplied their pet's name in 17/63 (27%) of the calls, while MS callers never did. MS callers did, on the other hand, occasionally volunteer information about the pet's breed or weight (as did RC callers).

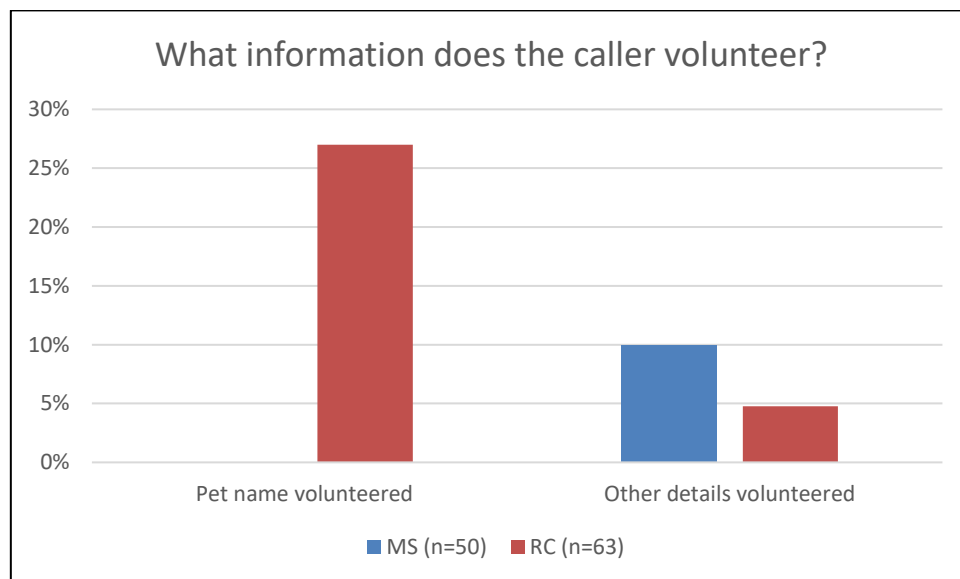


Figure 3. Proportion of calls, MS (blue) and RC (red), in which the caller's pet's name (left) or other information (e.g. breed, age, weight; right) was 'volunteered'. Proportion values are normalised according to respective totals n RC and MS calls.

Returning to the reference terms by which callers referred to their pet (e.g., "my dog"), while MS callers also used locally initial reference terms, and a majority of possessive pronouns, they took the form of an indefinite reference (e.g. "I've got a kitten") more often than in RC calls. Extracts 13-15 are examples.

Extract 13: MS Jabs 7

01 C: .hh Hiya, um I've got a kitten who needs he:r first lot of
 02 vaccinations.=I was just wondering >if y'could tell me< how
 03 much that would cost,

Extract 14: MS 2

01 C: Hi.=I [(got) a new d- uh: puppy the other day.
 02 V: [(coughs))
 03 C: .hh s'wonderin' how much it'd cost t'get the jabs done.
 04 please.

Extract 15: MS Jabs 2

01 C: .thhh hiya, Yeah.=I just ↑wanted to know how much you charge
 02 for uh puppy: vaccinations.

A chi-squared test (Figure 4) showed that the difference in use of indefinite article to refer to the object of request is unlikely to be random⁴, and that only RC callers refer to their pet by name.

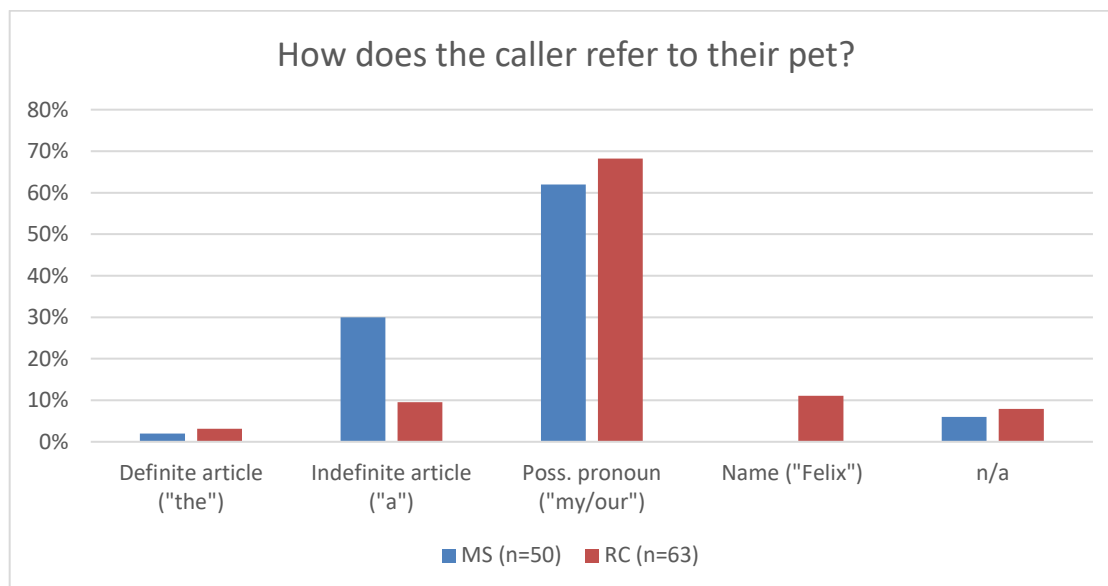


Figure 4. Reference forms to pets for MS (blue) and RC (red) calls. Proportion values are normalised according to respective totals n for RC and MS calls.

Close analysis revealed several turn design features of RC calls, absent in MS calls, that particularized the reality of having a pet, and displayed a stake in or commitment to it. We note that using a personal pronoun in a cost inquiry (e.g., “how much do you charge for MY puppy vaccinations”) would be an unlikely turn construction for native speakers. Differences are not absolute, however. Extracts 16-17 show that, while less frequent, RC callers both call to ask about cost and use the possessive pronoun to do so.

Extract 16: RC Vaccine 42

01 C: .mt Hi. Uhm I'm >just wondering how much it would be< to
02 get u:h some injections done for my puppy.

Extract 17: RC Vaccine 46

01 C: Yes good morning.=My name's Steve. I have ā:(ou-) our rabbit
02 Topsy is booked in for an appointment tomorrow afternoon;
03 .hhh for a follow up consult and a: a vaccination? .h[h and]
04 V: [Right,]
05 C: I was wondering if I could just confirm know how much
06 tomorrow's appointment is going to cost so that I can bring
07 along the right amount in cash.

In Extract 16, the caller asks for the price of injections, but on behalf of “my puppy” – a formulation we do not find in the MS dataset. In Extract 17, the caller seemingly starts a

⁴ ($X^2 = 7.725$, $df = 1$, $p < 0.01$)

formulation with the indefinite article ‘a’ (note, that the pronunciation cuts off ‘ay’ not the start of the word ‘an’, as in ‘an appointment’), repaired to “our rabbit Topsy”, and reference to “tomorrow’s appointment” before giving an account for asking this question: “so that I can bring along the right amount in cash.”

Based on these analysis, we suggest that these differences are accounted for by the stake RC callers have in the calls as well as the veracity of the questions they ask: RC callers (presumably) have authentic reasons for needing a vet appointment; MS callers make generalized requests for information about a pet they do not own. Real callers do so too, but more rarely. One relevant case is Extract 18. The caller makes reference to the pet with an indefinite article, but to a dog weighing “twenty-one kilos”.

Extract 18: RC Vaccine 73

01 V: How can I help you.
 02 (0.5)
 03 C: Hello, yes.=I wonder if you can.=Could you tell me how much
 04 an <annual ↓booster vaccination> would be for a dog weighing:
 05 twenty-one kilos. please.

Thus, the caller offers detail about her pet which is relevant to how vaccination charges are calculated. This enables the receptionist to respond immediately without having to solicit that information in a subsequent turn. Furthermore, unlike MS requests for vaccination costs, the caller specifies “annual booster vaccination” (line 04), in contrast to “the jabs” and “puppy vaccinations” (Extracts 14 and 15).

Another observation about how pets were referred to, in addition to the use of possessive pronoun or indefinite article, was the inclusion of hesitation or perturbation markers in their requests. It is, of course, a routine feature of real talk that it is littered with “uhs”, “ums”, repair initiators, pauses, and so on. Clark and Fox Tree (2002) describe uh(m) as a word, and one that projects upcoming trouble in speaking. Schegloff (2010) describes four sequential, turn-constructive, positioned uses of uh(m), including in repair operations. These include a search for a word, a replacement of a word in production, replacing one TCU in favour of another, and initiating ‘dispreferred’ first and second pair parts. In terms of repair, Schegloff writes that, “All of these “uh(m)”s occur somewhere in the course of dealing with some trouble in what has just been said or with something planned—and perhaps projected—to be said just up ahead, in that TCU, in that turn” (p. 138).

We found suggestive differences in the placement of these markers in RC versus MS calls. In Extracts 9-11, earlier, callers asked about “my dog”, “my lab Brownie” and “my cats”, and perturbations appeared in their accounts for appointment (“He’s had (.) (as) a bad ear.”; “he:’s (0.2) off his food”). Additionally, in Extract 10, the caller rang “regarding um my lab Brownie”. In Extract 11, the caller asked about an appointment “for: uh both my cats”. While RC callers do display category trouble in formulating their requests (e.g., in Extract 17 the caller refers to “ā:(ou-) our rabbit Topsy”), note that there are *no* perturbations, however, between the possessive pronoun “my” and the pet category.

In contrast, in Extract 14, above, the MS caller refers to a “new d- uh: puppy”, in which the perturbation is located *between* the indefinite article and pet category, and there is a further repair initiator as the caller presumably replaces ‘dog’ with ‘puppy’. Extract 19

provides another example of an inserted hesitation marker in the caller's first mention of the object of their request.

Extract 19: MS-1

01 C: Hi: I need to get my um dog castrated, I jus' wonder'd h'w
02 >much it'd< cost me.

A summary of the placement of “uh/ms” in caller requests is provided in Figure 5, below. MS callers stumble before the pet category in 18% of calls; RC callers in 10%. MS callers stumble more frequently ahead of the service category (28%) compared to RC callers (19%).

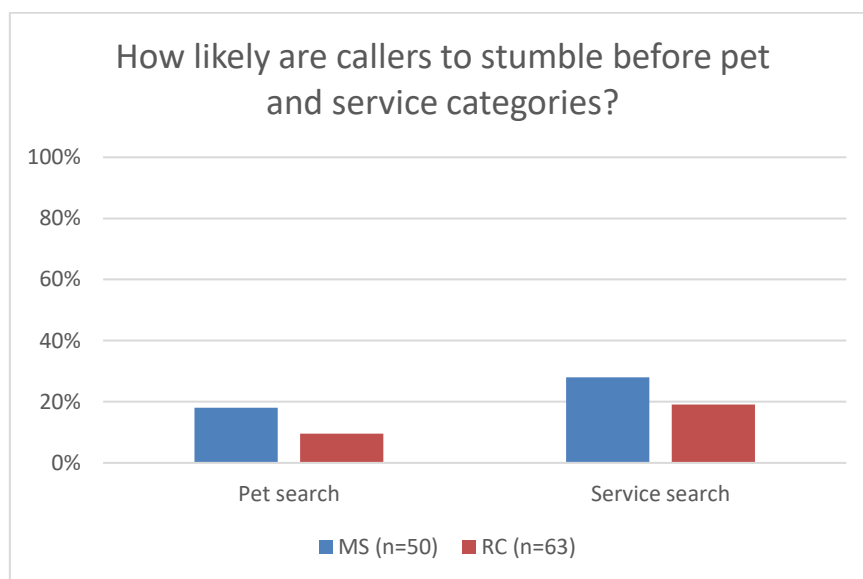


Figure 5. Placement of uh/ms in RC and MS datasets. Proportion values are normalised according to respective totals n for RC and MS calls.

These are relatively minor differences, but further analysis may reveal systematic differences in terms of which category of search (pet or service) occasions disfluent talk for RC and MS callers.

b. Responding to questions about pets

As noted above, RC and MS callers differed in the information they volunteered as part of their reason for the call. While RC callers regularly included their pet's name, for instance, in the design of their request, MS callers never did, at least in our dataset. This makes certain actions on the receptionist's behalf (in)apposite in the unfolding call. However, we also identified differences in the way callers responded to questions. In Extract 20, the receptionist asks the caller a question.

Extract 20: RC-Other 12

01 V: A:nd the cat's name?
02 C: Fluffy.

The receptionist asks for the pet's name, initiating their request with an and-preface, which connects their current action to an ongoing sequence (Heritage & Sorjonen, 2004). The RC caller responds immediately with a one-word TCU comprised of their pet's name.

Asking for a pet's name is relevant to booking an appointment for an animal. It is not surprising, then, that MS callers, asking about service cost, were not routinely asked for this information. However, Extract 21 provides a rare instance.

Extract 21: MS jabs

01 V: An' what've y'decided t'call 'im,
02 (0.2)
03 C: ↑(Yeh-/think) we're gonna call 'im Rufus.

In this case, the initiating action is different – it prefers the pet's name in response but is also about decision-making about that name. The MS caller's response is delayed by an inter-turn gap (line 02), turn-initial stumble (“↑(Yeh-/think)”), and a fully articulated answer (“we're gonna call 'im”); they do not just say “Rufus”.

While RC callers often included their pet's name and breed in initial requests, receptionists often asked about breed in calls about dogs. In Extract 18, the caller asked about “a dog weighing: twenty-one kilos”. Including information about weight marked it as an RC, rather than an MS, call. Extract 22 comes from the MS dataset.

Extract 22: MS 1

01 V: How: 1- much does your dog wei:gh?
02 (0.5)
03 C: Um: how much does 'e wei:gh, (0.3) oh I've no idea,

The MS caller responds to a routine vet question in ways that marks it as in unexpected or inapposite. Like Extract 21, the caller's response is delayed but only by an inter-turn gap at line 02, but also by repeating the question before marking the question as unexpected: the second TCU which responds to the question is initiated by a change-of-state token (Heritage, 1984): “Oh I've no idea”. Overall, the caller produces a dispreferred response – not supplying the requested information at all.

We found that out of 19 MS calls, 15 of the MS callers stumbled when producing responses about breed or weight, whereas in the two existing cases of this kind of question in our RC dataset, the caller was always fluent. More cases may be revealing, made evident through differences between RC and MS callers' delayed, dispreferred, or otherwise ‘stumbled’ production of their responses to questions about pets. However, any future coding of this kind would have to take into account how different ‘reasons for the call’ produce different subsequent trajectories and contingencies. For example, RC callers often supply their pet's name (and/or breed, weight), so receptionists are less likely to ask questions to elicit such information than in MS calls. Similarly, because most MS calls inquire about cost rather than booking an appointment, the name, breed, and weight of a pet is less relevant to fulfilling the request.

The implication of this result, and the preceding sections of analysis, is that RC callers know, and can anticipate, the information required about their pet in order to progress service at the vet's, while MS callers display a heavily constrained remit for their interactions.

DISCUSSION

The aim of this paper was to explore how human interlocutors accomplish the job of being an 'authentic' party to an interaction. By comparing the telephone inquiries of real customer (RC) callers to veterinary practices with simulated clients, or 'mystery shopper' (MS) callers, we found differences at both macro and micro levels of interaction. First, at the macro level, we identified some core differences between the two datasets. RC calls comprised more, and different, activities than MS calls. MS callers' 'reason for the call' was unvarying: requesting information about the cost of a service (e.g., vaccinations, euthanasia) or how a process worked. RC callers requested appointments for their pets, not information. Making appointments often involved sequence expansions into eliciting or supplying pet information, offering times and dates, comparing diaries, and so on. These structural differences in the interactional landscape of RC and MS calls led to other 'macro' differences, including that RC calls were simply longer than MS calls, and with more deviation around the average duration. Furthermore, the two datasets also differed in terms of call outcome, with RC calls typically ending in the confirmation of an appointment, and MS calls in 'no further action' – indicating the fact that RC callers had or were creating ongoing relationships with their surgeries, while MS callers resisted offers from receptionists. Finally, we found that RC callers 'reasons for the call' included actions that are entirely missing in the MS calls, such as accounting for requests in terms of sick pets for whom they had concerns or cancelling already-made appointments.

At the micro level, when analysing the precise components of turn design as MS and RC callers built their requests, we identified further differences. In addition to the purpose of the request (cost of service; making an appointment) RC callers included components that MS callers did not, such as accounts for their requests, or the name, breed, and weight of their pet. These differences made relevant or closed subsequent trajectories in talk. For instance, given that MS callers did not volunteer the name of their pet, receptionists sometimes asked. We observed that MS callers were more likely than RC callers to refer to their pets categorially, using the indefinite article to enquire about "a cat". Both MS and RC callers predominantly used a possessive pronoun to refer to their pet (e.g., "my cat"); however, only RC callers referred to their pet by name. We also presented findings that suggested MS callers were more likely than RC callers to initiate repair on, delay, or otherwise 'stumble' in responding to queries about the pet breed and weight.

Mystery shopping, as a research methodology, hinges on MS calls being functionally indistinguishable from real service calls. However, overall, we found that mystery shopper callers did not reproduce the social actions that comprise real customer telephone calls to veterinary practices. MS callers produced a limited set of 'scenarios' for the receptionists to progress, and that were generally atypical of RC callers.

The implications of our research for mystery shopping methodology, and the use of simulated clients more generally, are twofold. On the one hand, we have shown that, if mystery shoppers report back to organizations on the communicative competencies of call-takers, their report is based on a seriously constrained notion of their regular activity on the phone, and the range of contingencies they must deal with (e.g., cancelling appointments; sick animals). On the other, rather than dismiss the method entirely, this paper provides empirical grounds for improving the practice that market research organizations can act on. We argue that MS organizations should conduct research on the encounters that their real users or customers engage in, perhaps including making (and later cancelling) appointments.

Our findings also reinforce research that highlights the inauthenticity of role-played interaction and thus the value and integrity of training, and especially *assessing*, someone's 'communication skills' depending on what happens inside a simulated encounter (Stokoe, 2013; Stokoe & Sikveland, 2017). As Atkins et al (2016) point out, "the linguistic problems and differences that arise from interacting in artificial settings are of considerable importance in assessment, where we must be sure that the exam construct adequately embodies the skills expected for real-life practice". Our analysis shows how scrutiny of the ways people actually talk, and what they talk about, can reveal differences between MS and RC calls that would be difficult, if not impossible, to imagine. It also reveals that, and how, a person's stake in encounters matters (Edwards & Potter, 1992). For MS callers, who do not have a pet (or at least do not have the invented pet of their call), their stake is as an assessor of interaction, that they must remember and subsequently report on. For RC callers, their stake is in taking care of their pet, including medical interventions for sick and dying animals. To paraphrase De la Croix and Skelton (2009), the game of assessment overrides the game of pet care.

Finally, our findings have implications for the development of conversational AI systems such as Apple's *Siri*, Amazon's *Alexa*, and Google's *Duplex* (Leviathan & Matias, 2018) that attempt to emulate naturalistic spoken interaction. There is an emerging body of conversation analytic research that shows how constrained the 'conversations' are in such interactions (e.g., Albert et al, 2019; Reeves, 2017). Our findings suggest that, worse still, even *humans* who are simulating other humans do not produce social actions (e.g., requests) in the same way when the stakes are different. If humans attempting to simulate other humans cannot – at least from the evidence in this study – convincingly simulate social actions, the developers of artificial agents face an even tougher challenge. Before even approaching any engineering problems, they first have to solve the long-standing empirical question of how social actions are formed and recognized in interaction (Levinson, 2012). Without this crucial step, the systems they design will inevitably fail what we might call a 'Conversation Analytic Turing Test' (Albert et al, forthcoming), where the simulated interaction should yield functionally equivalent social actions and, thereby, similar social relationships. As simulated conversational agents proliferate and become increasingly central within service design, we stress the necessity for underpinning their development with empirical, conversation analytic studies of real talk.

Acknowledgements

We would like to thank Andrea Whittle and Catherine Woods for their very helpful comments while drafting this paper.

Biographies

Elizabeth Stokoe is Professor of Social Interaction in the School of Social Sciences at Loughborough University. Her research interests are in conversation analysis and membership categorization in various settings. She is the co-author of *Discourse and Identity* (with Bethan Benwell, Edinburgh University Press, 2006); co-editor of *Conversation and Gender* (with Susan Speer, Cambridge University Press, 2011), and author of *Talk: The Science of Conversation* (Little, Brown, 2018). She pioneered the training method called the Conversation Analytic Role-play Method (CARM).

Dr Rein Sikveland is a Research Associate in the School of Social Sciences at Loughborough University. His expertise is in conversation analysis and phonetics. Following his PhD (2011) at the University of York, Rein researched call centre interactions as part of his postdoctoral training at the University of Oxford. Dr Sikveland currently studies the interactional management in phone conversations between the public and commercial, health and mediation services. His research is currently applied to the Conversation Analytic Role-play Method.

Dr Saul Albert is Lecturer in Social Sciences at Loughborough University. Following his PhD (2017) in Cognitive Science at Queen Mary, University of London's Media and Arts Technology programme, he was a postdoctoral associate at the Human Interaction Lab at Tufts University (2016-2019). His research combines conversation analysis with participatory, computational, and practice-based methods in order to understand human interaction with a particular focus on aesthetics and judgments of taste as they are produced in everyday situations.

Dr Magnus Hamann is a Research Associate in the School of Social Sciences at Loughborough University. He completed his PhD (2018) at Aarhus University, focussing on memory and engagement in interactions with people with Traumatic Brain Injury (TBI). He has since worked as a postdoctoral researcher at the University of Southampton and also at Loughborough University working in the Conversation Analytic Role-play Method project in various settings, especially service encounters. Dr Hamann also is also a partner in a company that develops and facilitates linguistic workshops and teaching material for around 20.000 students each year.

William Housley is Professor of Sociology in the School of Social Sciences at Cardiff University. He is an expert in qualitative and social research methods, sociological theory, the study of practical reason, ethnomethodology, membership categorization analysis, social interaction and digital sociology. He was a co-founder and core member of the Collaborative Online Social Media ObServatory (COSMOS) and currently convenes the Digital Sociology

Research Group at Cardiff University. His contribution to Sociology was confirmed through the award of a DSc Econ. by Cardiff University in 2012 and was awarded the prestigious Vincent Wright Chair, at Sciences Po, Paris, for 2017.

References

- Albert, S.A., Housley, W., & Stokoe, E. (2019). In case of emergency, order pizza: An urgent case of action formation and recognition. *First International Conference on Conversational User Interfaces (CUI 2019)*, August 22-23, 2019, Dublin, Ireland. ACM, New York, NY, USA, <https://doi.org/10.1145/3342775.334280>
- Albert, S.A., Stokoe, E., Housley, W., & Sikveland, R.O. (frth.), *The Conversation Analytic Turing Test: Using conversation analysis to evaluate Google Duplex*.
- Atkins, S. (2018). Assessing health professionals' communication through role-play: An interactional analysis of simulated versus actual general practice consultations. *Discourse Studies*, 21 (2), 109-134. DOI: 10.1177/1461445618802659
- Atkins, S., Roberts, C., Hawthorne, K., & Greenhalgh, T. (2016). Simulated consultations: A sociolinguistic perspective. *BMC Medical Education*, 16 (16). DOI: 10.1186/s12909-016-0535-2.
- Bateson, J.E.G. (1992). *Managing services marketing* (2nd edition). Orlando, FL: Dryden Press.
- Bokken, L., Rethans, J.J., van Heurn, L., Duvivier, R., Scherpbier, A., & van der Vleuten, C. (2009). Students' views on the use of real patients and simulated patients in undergraduate medical education. *Academic Medicine*, 84 (7), 958-963.
- Calvert, P. (2005). It's a mystery: Mystery shopping in New Zealand's public libraries. *Library Review*, 54 (1), 24-35.
- Clark, H.H., & Fox Tree, J. (2002). Using uh and um in spontaneous speaking. *Cognition*, 84 (1), 73-111.
- Couper-Kuhlen, E. (2001). *Constructing reason-for-the-call turns in everyday telephone conversation*. Konstanz: Interaction and Linguistic Structures InLiSt No. 25
- De la Croix, A., & Skelton, J. (2009). The reality of role-play: Interruptions and amount of talk in simulated consultations. *Medical Education*, 43, 695-703.
- De Ruiter, J. P., & Albert, S. (2017). An appeal for a methodological fusion of conversation analysis and experimental psychology. *Research on Language and Social Interaction*, 50(1), 90-107.
- Edwards, D. & Potter, J. (1992). *Discursive psychology*. London: Sage.
- Fisher, N.I., & Kordupleski, R.E. (2018). Good and bad market research: A critical review of Net Promoter Score. *Applied Stochastic Models in Business and Industry*, 35 (1), 138-151.
- Garfinkel, H., (1967). *Studies in ethnomethodology*. Prentice Hall
- Glasier, A., Manners, R., Loudona, J.C., & Muir, A. (2010). Community pharmacists providing emergency contraception give little advice about future contraceptive use: A mystery shopper study. *Contraception*, 82 (6), 538-542.

- Goffman, E.A. (1974). *Frame analysis: An essay on the organization of experience*. Boston: Northeastern University Press.
- Hepburn, A. (2004). Crying: Notes on description, transcription, and interaction. *Research on Language and Social Interaction*, 37, 251-290.
- Hepburn, A., & Bolden, G. B. (2017). *Transcribing for social research*. London: Sage.
- Heritage, J. (1984). A change-of-state token and aspects of its sequential placement. In J.M. Atkinson & J. Heritage (Eds.), *Structures of social action: Studies in Conversation analysis* (pp. 299-345). Cambridge: Cambridge University Press.
- Heritage, J., & Sorjonen, M. (1994). Constituting and maintaining activities across sequences: And-prefacing as a feature of question design. *Language in Society*, 23, 1–29.
- Jacob, S., Schiffino, N., & Biard, B. (2018). The mystery shopper: A tool to measure public service delivery? *International Review of Administrative Sciences*, 84(1), 164–184.
- Leviathan, Y., & Matias, Y. (2018). *Google Duplex: An AI system for accomplishing real world tasks over the phone*. Retrieved from <https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html> on 2018-12-11
- Levinson, S. C. (2012). Action formation and ascription. In J. Sidnell & T. Stivers (Eds.), *The Handbook of Conversation Analysis* (pp. 101–130). Oxford: John Wiley & Sons.
- Mounsey, A.L., Bovbjerg, V., White, L., & Gazewood, J. (2006). Do students develop better motivational interviewing skills through role-play with standardised patients or with student colleagues? *Medical Education*, 40 (8), 775-780.
- Norris, P.T. (2002). Purchasing restricted medicines in New Zealand pharmacies: Results from a ‘mystery shopper’ study. *Pharmacy World & Science*, 24 (4), 149-153.
- Ott, B. (2016). The limits of control in service work: Interactive routines and interactional competence. *Research in the Sociology of Work* (Research in the Sociology of Work, Vol. 29, pp. 155-183). Bingley: Emerald Group Publishing Limited.
- Owen R. (2019). Net Promoter Score and its successful application. In K. Kompella (ed.), *Marketing wisdom: Management for Professionals*. Singapore: Springer.
- Pillet-Shore, D. (2017). Preference organization. In J.F. Nussbaum (Ed.), *Oxford research encyclopaedia of communication*. Oxford: Oxford University Press.
- Pilnick, A., Trusson, D., Beeke, S., O’Brien, R., Goldberg, S., & Harwood R.H. (2018). Using conversation analysis to inform role play and simulated interaction in communications skills training for healthcare professionals: Identifying avenues for further development through a scoping review. *BMC Medical Education*, 18 (1), 627. doi: 10.1186/s12909-018-1381-1.
- Reeves, S. (2017). Some conversational challenges of talking with machines. In *Talking with Conversational Agents in Collaborative Action*, Workshop at the 20th ACM conference on Computer-Supported Cooperative Work and Social Computing (CSCW ’17), Portland, Oregon, USA.
- Sacks, H. (1992). *Lectures on conversation*, Vol. I, edited by G. Jefferson. Oxford UK, Blackwell.
- Schegloff, E.A. (1986). The routine as achievement. *Human Studies*, 9, 111-151.
- Schegloff, E.A. (1996a). Confirming allusions: Toward an empirical account of action. *American Journal of Sociology*, 102 (1), 161-216.

- Schegloff, E.A. (1996b). Some practices for referring to persons in talk-in-interaction: A partial sketch of a systematics. In B. Fox (ed.), *Studies in anaphora* (pp. 437–485). Amsterdam: John Benjamins Publishing Company.
- Schegloff, E.A. (2010). Some other “Uh(m)”s’. *Discourse Processes*, 47 (2), 130–174.
- Schegloff, E.A., & Sacks, H. (1973). Opening up closings. *Semiotica*, 8 (4), 289–327.
- Stivers, T. (2012). Sequence organisation. In T. Stivers & J. Sidnell (Eds.), *The Handbook of Conversation Analysis* (pp. 191–209). Oxford: John Wiley & Sons.
- 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 conversation and the implications for communication training. *Research on Language and Social Interaction*, 46 (2), 1-21.
- Stokoe, E., & Sikveland, R.O. (2017). The Conversation Analytic Role-play Method: Simulation, endogenous impact and interactional nudges. In V. Fors, T. O’Dell & S. Pink (Eds.), *Theoretical scholarship and applied practice*. Oxford: Berghahn Books.
- Turing, A.M. (1948). Intelligent machinery: National Physical Laboratory Report. In B. Meltzer & D. Michie (Eds.), *Machine Intelligence 5*. Edinburgh: Edinburgh University Press.
- Van Hasselt, V.B., Romano, S.J., & Vecchi, G.M. (2008). Role playing: Applications in hostage and crisis negotiation skills training. *Behavior Modification*, 32 (2), 248-263.
- Wilson, A.M. (1998). The role of mystery shopping in the measurement of service performance. *Managing Service Quality: An International Journal*, 8(6), 414-420.

ⁱ Although Ott (2016) found, in his research on service providers’ experiences of using an organization’s standardized scripts for interacting with customers, that they are able to distinguish between mystery shoppers and ordinary service users. Staff reported that they “strategically adhere to and drop the [company-prescribed] script”, allowing “them the ability to appear to be following procedures when they think they are being graded, and to break the rules when they believe they are not under mystery shopper observation” (p. 174).