The Softness of Hard Data:

Discursive Psychology, Conversation Analysis, and Psychological Science

Elizabeth Stokoe[1], Charles Antaki[2], Leanne Chrisostomou[3],

Elle Henderson[4], and Simon Stewart[5]

[1] Department of Psychological and Behavioural Science, The London School of Economics and Political Science, UK

[2] Discourse and Rhetoric Group, Loughborough University, UK

[3v] School of Psychology, Sport and Health Sciences, University of Portsmouth, UK

[4] Open Polytechnic New Zealand

[5] Health, Education, Policing and Sciences, Staffordshire University, UK

Accepted for publication in *Qualitative Psychology*, 28.1.26

Abstract

Qualitative methods are sometimes criticised on the grounds that they do not provide 'hard' data. But, on inspection, 'hard' data turns out to be produced by unavoidably 'soft' human interaction and activities. That means that psychologists must work directly with what people do and say, and either transform it into abstractions – with potential distortions along the way – or stay with the raw events to see what questions they may answer. We argue for the latter: using discursive psychology and conversation analysis to ground claims about human sociality in the evidence that it provides, unfiltered, in everyday interaction. But, taking this argument further, we demonstrate how discursive psychological and conversation analytic scrutiny may reveal the 'softness' of both quantitative (experiments, standardisation) and qualitative (interview and survey questions) research tools, with implications for the production, openness, and validity of psychological knowledge.

Keywords: discursive psychology, conversation analysis, qualitative methods, data, research design, methodology

Public Significance Statement:

Research data that form the basis of psychological science rarely receive empirical scrutiny in terms of the way they are collected. The paper shows how the social interactions between researchers and their participants – in quantitative and qualitative research, from laboratory studies to interviews – impact the data collected and should be included in the open reporting of research.

Introduction

The origins of this paper and its title are in a series of exchanges between two epidemiologists and a mathematician on social media in 2021. In response to a report on Long Covid and children, one epidemiologist asked of the report, "Is there any data in there or is it just a list of anecdotes?" The other epidemiologist responded by saying "It's qualitative interview research: "111 children covering a range of topics related to effects of the pandemic." In response, the first epidemiologist wrote, "Thanks, this sounds like a great piece of science to me. I've got three children myself, … I'll ask each of them to submit a little paragraph." Finally, when taken to task by the mathematician for being "ill-informed about qualitative research", the first epidemiologist replied:

> I appreciate this was poorly worded and it doesn't really reflect my views on the merits of qualitative vs. quantitative science. My point was more whether we should give that much weight to qualitative studies on long-covid in children before we have collected any *hard data* (emphasis added).

This kind of bifurcation of research methods and data into 'hard' quantitative and 'soft' qualitative is well-established (e.g., Storer, 1967). If 'hard' data are the measurements, numbers, and statistics generated largely in laboratories, experiments, and surveys, then 'soft' data are words, texts, and images, generated largely in interviews, focus groups, and media. Evidence of the persistence of this bifurcation, its simplistic and erroneous nature, and its prejudicial manifestation in research evaluation can be found across academia, including in psychology (e.g., Clarke et al., 2024; Hedges, 1987; Land, 1981).

In this paper, however, we use qualitative methods to query the 'hard'-'soft' binary in terms of the production of research data themselves. We bring together three threads of discussion about psychological (social) science that flow from taking an interactional approach based in discursive psychology (DP) and conversation analysis (CA). The first thread locates DP and CA within the broader array of research methods in psychology. Both are, at a glance, qualitative methods that work with largely 'naturally occurring' social interactional materials: audio- and video-recordings and transcripts thereof. Yet, the underpinning theory and logic of DP and CA make them antithetical to the interview-based qualitative work, which dominates in the field, and thus hard to categorize. The second thread uses DP and CA to reflect on quantitative and qualitative research designs themselves. We argue that findings based on apparently 'hard' data may rest upon 'softer' foundations than previously thought, and that both quantitative and qualitative researchers often keep its *production* – that is, the 'soft' data of social interaction that comprises researcher-participant interactions across the research process – inside an uninterrogated black box (Stokoe, 2010; see also Masaryk & Stainton Rogers, 2024). The final thread considers what taking this interactional production seriously might mean for psychological science and how it, itself, yields 'hard' data. Thus, we consider what the practices that generate research data – 'hard' *or* 'soft' – mean for its collection, interpretation, and leverage.

Discursive Psychology, Conversation Analysis, and Qualitative Methodologies

Discursive psychology (DP) is the study of psychological matters in and for social interaction (Edwards, 1997; Potter, 1996). It is a research programme focused on how language accomplishes social action and constructs the world. More specifically, it investigates how people themselves deploy and orient to psychological concepts to do interactional work in everyday and professional settings (Stokoe et al., 2025). In a direct challenge to much of what we might regard as 'mainstream' psychology, Edwards and Potter (1992) pushed researchers to understand that, and how, language does things, rather than treating it as providing access to the mind. With its antecedents in the discourse and rhetorical approaches of the Loughborough group (e.g., Antaki, 1988; Edwards & Mercer, 1987; Billig, 1987; Potter & Wetherell, 1987), discursive psychologists promoted their blend of ordinary language philosophy, social studies of science, social constructionism, and ethnomethodology. Rather than treat language as window upon cornerstone psychological phenomena like attitudes, attributions, identities, personalities, emotion, or memory, via the laboratory or in questionnaires, surveys, focus groups, or interviews, they instead examined how language builds worlds and minds alike. In DP, the core purpose of language – in the widest sense across all its modalities – is the production and maintenance of intersubjectivity (see Stokoe, 2020).

Conversation analysis (CA) is a six-decades-old field of observational "cumulative science" (Stokoe, 2021, p. 348). It is both a method for capturing, transcribing, and analyzing naturally occurring social interaction and a theory of human sociality grounded in the systematic organization of social interaction which provides an infrastructure for social life and institutions (e.g., Heritage, 2009). Using audio- and video-recordings of naturally-occurring interaction 'in the wild,' CA uses a standardized technical system (Jefferson, 2004) for transcribing the produced details of talk and other conduct - words, phrases, grammar. non-lexical items, gaps, overlaps, gestures – which all combine to build, progress, and organize sequences of social interaction. That is, DP/CA collects, transcribes, and analyses data that pass what Potter (2002, p. 541) vividly referred to as the "dead social scientist test." That is, the focus is on capturing for analysis events that were not set up by or for research and thus that are not simulated, role-played, experimentally produced, or reported on post-hoc in interviews, focus groups, or surveys (Stokoe, 2013).

Discursive Psychology: Beyond The Quantitative/Qualitative Binary

While DP is typically categorized as a qualitative method, it is sometimes confused with other forms of qualitative discourse analysis and adaptations of what we take to be 'core' DP (i.e., the work flowing from Edwards & Potter, 1992), such as 'critical discursive psychology' (e.g., Parker, 2002; Wetherell, 1998). However, neither CA nor DP are usefully categorized as either quantitative or qualitative methods. In fact, qualitative psychologists have often been critical of DP/CA for its 'pointless empiricism', focusing on 'nothing but the text'; failing to address notions of subjectivity; being atheoretical and ahistorical; and thus, failing to address 'big and important questions' (summarized in Stokoe & Albert, 2024). These criticisms reflect similar characterizations of CA, including as "aggressively and programmatically

devoid of theoretical content of sociological relevance", and a method "in search of a substance" (Coser, 1975, p. 696). Such criticisms have always been refuted, of course, including through substantial empirical work focused on power, resistance, prejudice, and discrimination (for recent overviews, see Waring & Tadic, 2024; Whitehead et al., 2025).

One of the most common ways to differentiate 'hard' quantitative and 'soft' qualitative research is in the 'sample size' (e.g., the number of participants) or number of data points (e.g., the number of measures taken) and their generalizability (or 'transferability', see Drisko, 2025). What constitutes an appropriate participant sample to power an experiment or survey (e.g., Lakens, 2022) or 'saturate' an interview-based analysis (e.g., Czernek-Marszałek & McCabe, 2024) may comprise one of the ways to evaluate a study's rigour. However, research shows that reviewers often misapply criteria from quantitative to qualitative work (e.g., Clarke et al., 2024; Levitt et al., 2021) and that the criteria for one approach do not necessarily generalize to another (e.g., Cena et al., 2024; on quality criteria in DP/CA, see Antaki et al., 2003; Janusz & Peräkylä, 2020; O'Reilly et al., 2020). Furthermore, the issues of reliability and validity themselves comprise a vast and contested subfield in qualitative psychology (e.g., Makel et al., 2022; Parker, 2004; Pownall, 2025; Osbeck & Antczak, 2021).

This takes us to other reasons why DP is hard to categorize straightforwardly as hard/quantitative or soft/qualitative. One is that some DP/CA has included experimental methods and quantification in their research (e.g., De Ruiter & Albert, 2017; Potter, 2021) or leverage statistical methods to augment qualitative description of interactional phenomenon (e.g., Sikveland et al., 2016; Stivers, 2015), although this is not uncontroversial (e.g., Robinson, 2007; Schegloff, 1993). Indeed, the language and rhetoric of 'quantification' has itself been scrutinized by discursive psychologists (e.g., Billig, 2021; Potter et al., 1991) – acutely relevant in the rapid development of the discourse of 'data science' and 'big data' (e.g., Stevens et al., 2018). Another – perhaps more fundamental – reason is that what counts as a 'unit of analysis' (to be coded, counted, or otherwise) is not straightforward in DP/CA. For example, 'sample size' in both quantitative and qualitative psychology routinely refers to the numbers of research participants included in the study. However, in DP/CA, 'one' recorded interaction contains, for example, two or more participants and potentially hundreds of overlaps, pauses, gaps, 'ums', 'uhs', gestures, actions (e.g., question, answer, greeting, offer, request). This means that 'participants' are not counted, and nor are the 'numbers of conversations.' It is any one of these phenomena and their organization that may form units of analysis. As Aldrich (2014) points out, "methods of data collection do not automatically determine methods of data analysis." Another reason is that a key aim of much DP/CA research is to identify "repetitive, uniform, typical and *cohort-independent*" practices (Heap, 1990, p. 46, emphasis added). Such practices are often, as it turns out, 'generalizable' across settings, languages, and timelines, as identified across cumulative bodies of research papers (e.g., the practice of 'doing being ordinary': Sacks, 1984; see Lawrence, 1996; Li, 2025; Rowley, 2007; Wooffitt, 1991; Tovares, 2019).

Furthermore, DP/CA are forms of logical rather than numerical analysis (Stokoe, 2018). An example from linguistics makes the point[1]: while the sentence "Peter forwarded the letter to his aunt Mary" is grammatical, "to aunt the forwarded Peter letter Mary his" - which

[1] With thanks to Derek Edwards for providing this illustration.

contains the same words - is not. But understanding the sentence, either by speaker or analyst, is not a matter of counting *how many times* people say these things or asking how many people agree about their grammaticality. It is a matter of knowing how to speak English. DP/CA does not need large samples or probabilistic statistics to do an analysis, even though it does make collections of instances of phenomena to figure out how they work (for more on building collections and other related methodological practices and principles in CA, see Robinson et al., 2024). But even then, the analysis of each "instance" is done on the same basis, including recognizing something *as* an instance, which is the tacit ability that people have, and that anthropologists need, to understand the uses of their own natural language, or one in which they have acquired some competence. This is because the methodic order of interaction is manifest not only in aggregate but on each of its occasions. Schegloff (1992, p. xlvi, emphasis added) explains this with regards to Harvey Sacks's (the founder of CA) view of sampling:

> Taking up the methodological relevance of sampling, Sacks points out that it depends on the sort of order one takes it that the social world exhibits. An alternative to the possibility that *order manifests itself at an aggregate level and is statistical in character* is what he terms the 'order at all points' view (lecture 33, p. 484). This view, rather like the 'holographic' model of information distribution, understands order not to be present only at aggregate levels and therefore subject to an overall differential distribution, but to be present in detail on a case by case, environment by environment basis.

For this reason, DP/CA is also antithetical to the usual psychological sense of replication: there is no such thing as an interactional component being an outlier and thereby excludable from a dataset – participants cannot be 'wrong' (Stokoe et al., 2025). Instead, 'deviant cases' that apparently depart from patterns observed in 'typical' cases are examined for how they came to be produced as such by the participants, and what this may reveal about the broader patterns from which they appear to depart (see, e.g., Schegloff, 1968).

Having summarized the basic approaches of DP/CA, we now consider their implications for the collection of 'hard' (e.g., in the laboratory) and 'soft' (e.g., in interviews) data.

The Softness of Hard Data

One of the best-known psychology experiments is Milgram's (1963, p. 371) "procedure for the study of destructive obedience in the laboratory." The laboratory study involved the research participants delivering (ostensibly, but not actually) mild to life-threatening electric shocks to another participant (who was in fact Milgram's confederate) under standardized instructions of the experimenter. Milgram's conclusion was that people possess "two deeply ingrained behavior dispositions: first, the disposition not to harm other people, and second, the tendency to obey those whom we perceive to be legitimate authorities" (p. 378). Milgram's classic obedience experiments have generated much debate since their publication, often focused on ethics, but since then, some of the most important evaluations have come from DP/CA, including extensive analyses of the experiment's original audio-recordings.

For example, using conversation analysis to shed light on "the concrete, empirical details of what his participants actually said and did," Hollander (2015, p. 426; see also Hollander & Turowetz, 2017) showed that Milgram's categorization of participants into either 'obedient' or 'defiant' failed to focus on the continuum of practices through which participants resisted the experimenter's instructions. Instead, Hollander identified "a more detailed and nuanced picture of patterned action" than the binary categorization implied (p. 438), thus undermining (though not toppling) the much-perpetuated 'hard' fact that "65% of participants [obeyed] the highest level of 450 volts" (McLeod, 2025). Similarly, using DP, Gibson (e.g., 2013, 2019) found that, rather than deliver verbal instructions in replicable ways, "participants could draw the experimenter into a process of negotiation over the continuation of the experimental session" as well as "radical departures from the standardized experimental procedure" (2013, p. 290). In a conversation analytic study of experimenter-participant interaction in parapsychology laboratories, Wooffitt (2007, p. 477) also found that the way the "experimenter acknowledges the research participants' utterances may be significant for the trajectory of the experiment." Even apparently standardized questions for gaining research consent from participants, both on written forms and in spoken interaction, have been shown to tilt respondents towards a 'yes' (Speer & Stokoe, 2014).

The research above has clear implications for the integrity of 'hard' data (that flow from experiments and statistical summaries) and their epistemic upshots (that flow from failing to pay attention to the interactional and situated production of (social) science). Discussions of such issues are generally rare in psychology, despite there being an entire field dedicated to the study of the production of scientific knowledge (e.g., Gilbert & Mulkay,1984; Latour & Woolgar, 1979) and its 'construction yard' (Potter, 1996). When the interactions in and through which (social) science occurs become the topic of and not just a resource for the production of knowledge (Wieder, 1988) the 'hardness' of data begins to soften. As Dingemanse et al. (2023, p. 1) point out, "[t]hat social interaction matters is recognized by any experimentalist who seeks to exclude its influence by studying individuals in isolation."

One reason why 'hard' data collection is inevitably impacted by its interactional production is that our encounters are shaped by what conversation analysts call 'recipient design' (e.g., Houtkoop-Steenstra, 2000). Recipient design refers to the way people shape what they say word for word for the person they are talking to. This means that we are constantly editing, elaborating, shortening, and amending everything we say for the person we are talking to – often very subtly and without being aware in the millisecond pace of interaction. A kind of interactional imperative, therefore, shapes how we design turns at talk for our interlocutors, under the empirical norms of 'preference organization' (e.g., Pillet-Shore, 2017), 'progressivity' (e.g., Stivers, 2008), and other organizational features of conversation. Such phenomena may account for what happened inside Milgram's laboratory, given that we are only scratching the surface of the 'laboratory life' of psychologists.

Furthermore, in a series of conversation analytic studies of the verbal delivery of text-based survey questions, the researchers found that "mechanical objectivity" and standardization (Lavin & Maynard, 2001, p. 472) were assumed but not demonstrated to be present, through a range of practices including the reformulating and re-ordering of questions. Given the extent of research data collected via surveys, it is important to know not only how

the written design of each item may subtly constrain or afford responses, but also what happens when they are delivered verbally within social interaction. Heritage (2002) showed how the items on a text-based checklist are formulated into spoken talk by health visitors when attending new parents. For instance, items on the checklist like "Pregnancy normal/abnormal, specify" became "and you had a normal pregnancy" in a series of recipient-designed questions, which are grammatically and sequentially tilted to receive a 'no-problem' response ("yeh"). Wilkinson (2011, p. 358) similarly showed the tension between standardized instructions and their spoken delivery, and how responses get "transformed into entries on a coding sheet." Taking the case of an organization's ethnic monitoring reporting, she examined how official statistics were collected, including the fact that the call-takers ended up categorizing 86% of callers as 'White European' even though such a category was not included on the report itself.

The impact of small deviations in the delivery of so-called standardized items is consequential in other ways. For example, Antaki and Rapley (1996, p. 421) examined recordings of how a 'Quality of Life' diagnostic instrument was administered by psychologists to service-users. The instructions for "reading the items" included paying "close attention to the exact wording." Although each question had three response options, psychologists often reformulated them into one 'yes/no' question, one of several practices "through which interactional outcomes are promoted or favoured" (Clayman, 2002, p. 230). Since treatment and support options were tied to questionnaire scores, the re-doing of questions had important consequences for those being assessed. The researchers concluded that it is hard to "draw conclusions from simple aggregation of recorded responses to this questionnaire, and, perhaps, to any questionnaire using a fixed-response schedule" (Antaki & Rapley, 1996, p. 421).

Other researchers have identified similar problems with the verbal delivery of diagnostic instruments. For example, Jones et al. (2020) examined the delivery of a test (Addenbrooke's Cognitive Examination, 2012) to identify cognitive impairment. They showed that neurologists often deviate from the standardized procedures which are designed to ensure test accuracy and consistency, calling into question the validity of test outcomes. Similarly, Ford et al.'s (2020) analysis of the verbal delivery of the Patient Health Questionnaire (PHQ-9) depression tool, and Turowetz's (2015) analysis of how clinicians use the Psycho-Educational Profile (PEP) to diagnose autism, both revealed how interactional processes impact the production of apparently 'hard' clinical outcomes. And even when text-based instruments are enforceable in law, interactional forces shift interlocutors away from what is written. Research has found that, for instance, police officers deviate from what is mandated by the Police and Criminal Evidence Act (1984) when it comes to opening interviews (Stokoe, 2013); that the right to remain silent can be undermined by the moral and interactional imperative to respond to questions and fill silence (Stokoe et al., 2016); that questions in evidence-gathering interviews with vulnerable adults and victims deviate from written guidance (e.g., Richardson et al., 2019), and that narrative descriptions of events in the world get transformed and squeezed into some of the 'hardest' data categories around: whether or not someone broke the law (e.g., Stokoe & Edwards, 2008).

The implication of treating the production of research data as topic rather than resource is not limited to quantitative data. As noted previously, some of the strongest

objections to DP/CA come from other qualitative researchers. This may be in part due to the fact that the overwhelming majority of qualitative research in psychology (Dunwoodie et al., 2023) collects data using semi-structured interviews (or focus groups) which have been the target of critique in DP/CA. What (much) qualitative and quantitative research share is the extrapolation of what happens in a laboratory, survey, or interview to the naturally occurring setting of their lives. For instance, rather than study domestic relationships as they ebb and flow at home (e.g., Goodwin, 2007), researchers ask people about their relationships post-hoc (e.g., Beitin, 2008) – leading to debate about the artefactual production of knowledge. Potter and Hepburn (2005, p. 281) list the following problems in the reporting and analysis of semi-structured interviews:

> Contingent problems in the reporting of interviews include: (1) the deletion of the interviewer; (2) the conventions for representing interaction; (3) the specificity of analytic observations; (4) the unavailability of the interview set-up; (5) the failure to consider interviews as interaction. Necessary problems include: (1) the flooding of the interview with social science agendas and categories; (2) the complex and varying footing positions of interviewer and interviewee; (3) the orientations to stake and interest on the part of the interviewer and interviewee; (4) the reproduction of cognitivism.

Thus, while qualitative researchers might argue that interviews produce 'rich' data that adds much value to surveys, the interview situation "plays havoc with the motive force of the telling … by making the elicitation itself the invariant occasion for telling the story" (Schegloff, 1997, p. 99-100). Harré (1979, p. 115) makes a related point about "the use of questionnaires with limited range questions" which "produce[s] not a representation of the social world being studied, but the representation of the shadow cast upon the social world by the prior conceptual apparatus deployed by the person who constructed the questionnaire"

The points laid out above show that it is risky to assume that researchers 'just ask' and participants 'just respond' to any given question. One of the most powerful demonstrations of the impact of question design comes from within the core of psychology itself. In the 1970s, Loftus and colleagues, in their studies of eyewitness reports in experimental settings, demonstrated how the wording of questions about the events being witnessed influences responses. Loftus and Palmer (1974) showed participants films of car accidents and, in a first experiment, found that estimates of speed varied according to whether they were asked how it 'bumped', 'smashed', 'hit', 'collided', etc., the other. The 'smashed' condition elicited higher estimates from participants. Additionally, in a second experiment, when asked "Did you see any broken glass", participants were more likely to respond affirmatively if they had been asked about a 'smash' in the first experiment despite there being none in the film. Thus, "very subtle changes in wording influences a witness's report" (Loftus, 1975, p. 652). Even swapping definite for indefinite articles in question design made a difference (e.g., "did you see *the* broken headlight" versus "did you see *a* broken headlight").

A third iteration of Loftus's experiment might usefully draw on one of the most compelling examples of how question design impacts data collection. It comes from a randomized controlled trial in primary care that drew on conversation analytic research on

'yes/no' questions (e.g., Raymond, 2003). Heritage et al. (2007) compared patients' responses to questions asking, "is there anything else…?" with those asking, "is there something else…?" at the start of consultations. While both are interrogatives, the negatively polarized 'anything' version was more likely to get a 'no' in response, while the positively polarized 'something' version was more likely to get a 'yes.' The 'something' inquiry frame significantly increased the number of elaborated responses given by patients and reduced the number of unmet concerns in consultations. Returning to Loftus, then, we might speculate that comparing the original "Did you see any broken glass" with "Did you see some broken glass", might have yielded further evidence about the importance of each word in a question.

The way researchers (whether qualitative or quantitative) ask questions impacts the data collected, and therefore the content and validity of their findings. The difference between using "the" or "a", and "some" or "any" force our attention towards the 'hard' empirical reality that is revealed through interaction analysis. In other words, perhaps 'soft' data presents a 'hard' test. We consider this in the next section.

The Hardness of Soft Data

We mentioned earlier that DP/CA is not readily categorizable as either a qualitative or quantitative method – it is neither and both. As Stivers (2015, p. 3) notes:

> [F]or all its qualitative refinement, CA is arguably the most quantitative of the qualitative social science methods. This is true in two key respects. First, the categorizing of interactional phenomena requires a clear characterization of what "counts" as an instance of the target phenomenon. In this way the focus is not on the uniqueness of each and every snowflake of data but on the properties that these snowflakes share. Second, distributional evidence is a key part of the CA method. Specifically, CA studies rely on the broad patterning of properties in datasets. Neither of these features is present as constitutive features of methods that rely on thematic analyses of participant observation, interviews, or focus groups.

Despite the 'hard' power of these 'soft' methods, DP/CA is rarely part of undergraduate or postgraduate training. While teaching students "from the first day of class that psychology is the science of behavior and that its ultimate goal is to describe and explain what people do," psychology "pays remarkably little attention to the important things that people do" – instead using "introspective self-reports, hypothetical scenarios, and questionnaire ratings" (Baumeister et al., 2007, p. 396). Ten years after Baumeister et al.'s critique, Mehl (2017, p. 184) similarly argued that

> [l]aypersons often think of psychologists as professional people watchers. It is ironic, then, that naturalistic observation, as a methodology, has a remarkably thin history in our field. … At the same time, there are clear limitations to what self-reports can assess. … the psychological scientist's tool kit also needs a method to directly observe human behavior in daily life. … naturalistic observation can bring behavioral data collection to where moment-to-moment behavior naturally happens.

It is frustrating that neither of these psychologists mentions DP/CA as a corrective for the situation they critique, especially given the focus on how "ordinary activities get done methodically and *reproducibly*" (Schegloff, 1992, p. xvii, emphasis added). We argue that DP/CA methods pose interesting questions for the psychological bedrocks of reliability and validity, which are often presumed rather than demonstrated in, say, the method sections of research reporting. One of the benefits of DP/CA research is that our data are perhaps some of the most transparently presented. Since in DP/CA "recordings give the researcher direct access to the details of social action, and they make it possible to scrutinize the data over and over again" (Janusz & Peräkylä, 2021, p. 428), a key quality indicator is whether a sufficient stretch of interaction is provided so that someone evaluating or reviewing the work has enough information to evaluate it. Recently, the relevance of DP/CA's practices of data sessions, data sharing across the community, building open datasets, and discussing this in the context of ethics and participant anonymity, have figured in wider discussions about open (qualitative) science (see; Huma & Joyce, 2023; cf., Pownall, 2025). Perhaps, ironically for research methods that might seem grounded in the softest of all data – the 'mess' of naturally-occurring social interaction – our materials provide hard tests for the production of hard evidence about human action.

Conclusion

> Psychologists long abandoned the due philosophical, linguistic, and semantic work of describing their subject matter and thus building the conceptual framework defining it, despite historically good examples. Consequently, the discipline has become a scattered crowd of dazzled conductors pretending to decipher a deep mystery despite having only superficial, if not defective, decoders (Düzen & Uysal, 2024, p. 146)

This paper started with a strongly-worded exchange between two epidemiologists and a mathematician. One response to "whether we should give that much weight to qualitative studies on long-covid in children before we have collected any hard data" is to reverse the direction of this question to ask "how much weight we should give *quantitative* studies that relate to complex real-world processes until we understand something of the ways in which the data were collected" (Stokoe et al., 2021, emphasis in original). As Schegloff (1993, p. 114) wrote, "quantification is no substitute for analysis." We have given examples of that, and how, the production of 'hard' data – from how fast a car was travelling to how many 'yes' and 'no' responses – involves potential points of variation with respect to interaction and context that are only visible through collecting and analyzing the 'soft' encounters through which they are produced. We argue that it is the duty of psychological researchers across all research designs to study and be critically reflexive about the operation of their own methods. As noted above, central to these discussions are recent statements and guidance about open research, including in qualitative (e.g., Branney et al., 2023) and discursive psychological and conversation analytic (DP/CA, e.g., Albert et al., 2025; Huma & Joyce, 2023) work. While open (meta)data is a core tenet of open research, it is often people's interactional practices that are formulated *as* data, findings, and knowledge and

these data, are, in turn, obscured in the research process and from subsequent reporting. Indeed, perhaps the replicability crisis is nothing more than a failure to examine all aspects – from participant recruitment and ethics to the interactional but unreported black boxes of psychological research design.

References

Albert, S., Hoey, E. M., & Raymond, C. W. (2025). The case for open conversation analysis data. *Research on Language and Social Interaction*, *58*(2), 156-164. https://doi.org/10.1080/08351813.2025.2484992

Aldrich, H.E. (2014). Stand up and be counted: Why social science should stop using the qualitative/quantitative dichotomy. https://blogs.lse.ac.uk/impactofsocialsciences/2014/11/28/stand-up-and-be-counted-social-science-qualitative-quantitative-dichotomy/

Antaki, C. (Ed.). (1988). *Analysing everyday explanation: A casebook of methods.* Sage Publications, Inc.

Antaki, C., Billig, M., Edwards, D., & Potter, J. (2003). Discourse analysis means doing analysis: A critique of six analytic shortcomings. *Discourse Analysis Online*, *1*. https://hdl.handle.net/2134/633

Antaki, C., & Rapley, M. (1996). Questions and answers to psychological assessment schedules: hidden troubles in 'quality of life' interviews. *Journal of Intellectual Disability Research*, *40*(5), 421-437. https://doi.org/10.1046/j.1365-2788.1996.779779.x

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: whatever happened to actual behavior? *Perspectives on Psychological Science*, *2*(4), 396–403. https://doi.org/10.1111/j.1745-6916.2007.00051.x

Beitin, B. K. (2008). Qualitative research in marriage and family therapy: Who is in the interview? *Contemporary Family Therapy*, *30*, 48-58. https://doi.org/10.1007/s10591-007-9054-y

Billig, M. (2021). Rhetorical uses of precise numbers and semi-magical round numbers in political discourse about COVID-19: Examples from the government of the United Kingdom. *Discourse & Society*, *32*(5), 542-558. https://doi.org/10.1177/095792652110131

Billig, M. (1987). *Arguing and thinking*: *A rhetorical approach to social psychology.* Cambridge University Press.

Branney, P. E., Brooks, J., Kilby, L., Newman, K., Norris, E., Pownall, M., ... & Whitaker, C. M. (2023). Three steps to open science for qualitative research in psychology. *Social and Personality Psychology Compass*, *17*(4), e12728. https://doi.org/10.1111/spc3.12728

Cena, E., Brooks, J., Day, W., Goodman, S., Rousaki, A., Ruby-Granger, V., & Seymour-Smith, S. (2024). Quality criteria: General and specific guidelines for qualitative approaches in psychology research. A concise guide for novice researchers and reviewers. *International Journal of Qualitative Methods*, *23*, https://doi.org/10.1177/160940692412828

Clarke, V., Braun, V., Adams, J., Callaghan, J. E., LaMarre, A., & Semlyen, J. (2024). "Being really confidently wrong": Qualitative researchers' experiences of methodologically incongruent peer review feedback. *Qualitative Psychology*, *12*(1), 7–24. https://doi.org/10.1037/qup0000322

Clayman, S. E. (2002). Sequence and solidarity. In E. J. Lawler & S. R. Thye (Eds.), *Advances ingroup processes: Group cohesion, trust, and solidarity* (pp. 229–253). Elsevier Science.

Coser, L. (1975). Two methods in search of a substance. Presidential Address, Annual meetings of the ASA, 1974. *American Sociological Review, 40* (6), 691–700. https://doi.org/10.2307/2094174

Czernek-Marszałek, K., & McCabe, S. (2024). Sampling in qualitative interview research: criteria, considerations and guidelines for success. *Annals of Tourism Research*, *104*, 103711. https://doi.org/10.1016/j.annals.2023.103711

Dingemanse, M., Liesenfeld, A., Rasenberg, M., Albert, S., Ameka, F. K., Birhane, A., ... & Wiltschko, M. (2023). Beyond single-mindedness: A figure-ground reversal for the cognitive sciences. *Cognitive Science*, *47*(1), e13230. https://doi.org/10.1111/cogs.13230

Dingemanse, M., Roberts, S. G., Baranova, J., Blythe, J., Drew, P., Floyd, S., Gisladottir, R. S., Kendrick, K. H., Levinson, S. C., & Manrique, E. (2015). Universal principles in the repair of communication problems. *PLOS One*, *10*(9), e0136100. https://doi.org/10.1371/journal.pone.0136100

Drisko, J. W. (2025). Transferability and generalization in qualitative research. *Research on Social Work Practice*, *35*(1), 102-110. https://doi.org/10.1177/104973152412565

Dunwoodie, K., Macaulay, L., & Newman, A. (2023). Qualitative interviewing in the field of work and organisational psychology: Benefits, challenges and guidelines for researchers and reviewers. *Applied Psychology*, *72*(2), 863–889. https://doi.org/10.1111/apps.12414

Düzen, N. E., & Uysal, M. S. (2024). Placing semiotic formulations to the epistemological ground of psychology. *Annual Review of Critical Psychology*, *18*, 144-165. https://discourseunit.com/wp-content/uploads/2024/10/0144_duzen_uysal.pdf

Edwards, D. (1997). *Discourse and cognition*. Sage.

Edwards, D., & Mercer, N. (1987). *Common knowledge: The development of understanding in the classroom.* Methuen.

Edwards, D., & Potter, J. (1992). *Discursive psychology*. Sage.

Ford, J., Thomas, F., Byng, R., & McCabe, R. (2020). Use of the Patient Health Questionnaire (PHQ-9) in practice: Interactions between patients and physicians. *Qualitative Health Research*, *30*(13), 2146–2159. https://doi.org/10.1177/1049732320924625

Gibson, S. (2019). *Arguing, obeying and defying: A rhetorical perspective on Stanley Milgram's obedience experiments*. Cambridge University Press.

Gibson, S. (2013). Milgram's obedience experiments: A rhetorical analysis. *British Journal of Social Psychology*, *52*(2), 290-309. https://doi.org/10.1111/j.2044-8309.2011.02070.x

Gilbert, G. N. & Mulkay, M. (1984). *Opening Pandora's box: A sociological analysis of scientists' discourse*. Cambridge University Press.

Goodwin, M.H. (2007). Occasioned knowledge exploration in family interaction. *Discourse and* Society, *18*(1), 93-110. https://doi.org/10.1177/0957926507069459

Harré, R. (1979). *Social being*: *A theory for social psychology*. Basil Blackwell.

Heap, J. L. (1990). Applied ethnomethodology: Looking for the local rationality of reading activities. *Human Studies, 13,* 38-72. http://dx.doi.org/10.1007/BF00143040

Hedges, L. V. (1987). How hard is hard science, how soft is soft science? The empirical cumulativeness of research. *American Psychologist*, *42*(5), 443-455. https://doi.org/10.1037/0003-066X.42.5.443

Heritage, J. (2009). Conversation analysis as social theory. In B. S. Turner (Ed.), *The new Blackwell companion to social theory* (pp. 300-320). Blackwell.

Heritage, J. (2002). Ad hoc inquiries: two preference in the design of routine questions in an open context. In Douglas W. Maynard, Hanneke Houtkoop-Steenstra, Nora Cate Schaeffer, Johannes van der Zouwen (eds.), *Standardization and Tacit Knowledge: Interaction and Practice in the Survey Interview*. Wiley.

Heritage, J., Robinson, J. D., Elliott, M. N., Beckett, M., & Wilkes, M. (2007). Reducing patients' unmet concerns in primary care: the difference one word can make. *Journal of General Internal Medicine*, *22*(10), 1429–1433. https://doi.org/10.1007/s11606-007-0279-0

Hollander, M. M. (2015). The repertoire of resistance: Non-compliance with directives in Milgram's 'obedience' experiments. *British Journal of Social Psychology*, *54*(3), 425-444. https://doi.org/10.1111/bjso.12099

Hollander, M. M., & Turowetz, J. (2017). Normalizing trust: Participants' immediately post-hoc explanations of behaviour in Milgram's 'obedience' experiments. *British journal of Social Psychology*, *56*(4), 655–674. https://doi.org/10.1111/bjso.12206

Houtkoop-Steenstra, H. (2000). *Interaction and the standardized survey interview: The living questionnaire*. Cambridge University Press.

Huma, B., & Joyce, J. B. (2023). 'One size doesn't fit all': Lessons from interaction analysis on tailoring Open Science practices to qualitative research. *British Journal of Social Psychology*, *62*(4), 1590-1604. https://doi.org/10.1111/bjso.12568

Janusz, B., & Peräkylä, A. (2020). Quality in conversation analysis and interpersonal process recall. *Qualitative Research in Psychology*, *18*(3), 426–449. https://doi.org/10.1080/14780887.2020.1780356

Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In G.H. Lerner (Ed.), *Conversation analysis: Studies from the first generation* (pp. 13-31). John Benjamins Publishing Company.

Jones, D., Wilkinson, R., Jackson, C., & Drew, P. (2020). Variation and interactional non-standardization in neuropsychological tests: The case of the Addenbrooke's Cognitive Examination. *Qualitative Health Research*, *30*(3), 458–470. https://doi.org/10.1177/1049732319873052

Lakens, D. (2022). Sample size justification. *Collabra: Psychology*, *8*(1), 33267. https://doi.org/10.1525/collabra.33267

Land, K. C. (1981). On the use of "hard" and "soft" methodologies in sociology. *The Sociological Quarterly*, *22*(3), 452-458. https://www.jstor.org/stable/4106129

Latour, B. & Woolgar, S. (1979). *Laboratory life: The construction of scientific facts*. Sage.

Lavin, D., & Maynard, D. W. (2001). Standardization vs. rapport: Respondent laughter and interviewer reaction during telephone surveys. *American Sociological Review*, *66*(3), 453-479. https://doi.org/10.2307/3088888

Lawrence, S. G. (1996). Normalizing stigmatized practices: Achieving co-membership by doing being ordinary. *Research on Language and Social Interaction*, *29*(3), 181-218. https://doi.org/10.1207/s15327973rlsi2903_1

Levitt, H. M., Morrill, Z., Collins, K. M., & Rizo, J. L. (2021). The methodological integrity of critical qualitative research: Principles to support design and research review. *Journal of Counseling Psychology*, *68*(3), 357-360. https://doi.org/10.1037/cou0000523

Li, X. M. (2025). Doing being an average teenager: Deploying ordinariness as subversive disability performance in presentational media. *Language in Society*, *54*(3), 439-464. doi:10.1017/S0047404524000368

Loftus, E.F. (1975). Leading questions and the eyewitness report. *Cognitive Psychology*, *7*, 560-572. https://doi.org/10.1016/0010-0285(75)90023-7

Loftus, E.F., & Palmer, J.C. (1974). Reconstruction of auto-mobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior*, *13*, 585-589. https://doi.org/10.1016/S0022-5371(74)80011-3

Makel, M. C., Meyer, M. S., Simonsen, M. A., Roberts, A. M., & Plucker, J. A. (2022). Replication is relevant to qualitative research. *Educational Research and Evaluation*, *27*(1–2), 215–219. https://doi.org/10.1080/13803611.2021.2022310

Masaryk, R., & Stainton Rogers, W. (2024). The time has come for psychology to stop treating qualitative data as an embarrassing secret. *Social and Personality Psychology Compass*, *18*(2), e12938. https://doi.org/10.1111/spc3.12938

McLeod, S. (2025). Stanley Milgram shock experiment. *Simply Psychology* https://www.simplypsychology.org/milgram.html

Mehl, M. R. (2017). The Electronically Activated Recorder (EAR). A method for the naturalistic observation of daily social behavior. *Current Directions in Psychological Science*, *26*(2), 184–190. https://doi.org/10.1177/0963721416680611

Milgram, S. (1963). Behavioral study of obedience. *The Journal of Abnormal and Social Psychology, 67*(4), 371–378. https://doi.org/10.1037/h0040525

O'Reilly, M., Kiyimba, N., Lester, J. N., & Edwards, D. (2020). Establishing quality in discursive psychology: Three domains to consider. *Qualitative Research in Psychology*, *18*(3), 406–425. https://doi.org/10.1080/14780887.2020.1810373

Osbeck, L. M., & Antczak, S. L. (2021). Generalizability and qualitative research: A new look at an ongoing controversy. *Qualitative Psychology*, *8*(1), 62-68. https://doi.org/10.1037/qup0000194

Parker, I. (2004). Criteria for qualitative research in psychology. *Qualitative Research in Psychology*, *1*(2), 95-106. https://doi.org/10.1191/1478088704qp010oa

Parker, I. (2002). *Critical discursive psychology.* Palgrave Macmillan.

Pillet-Shore, D. M. (2017). Preference organization. In J. Nussbaum (Ed.), *Oxford Research Encyclopedia of Communication*. Oxford University Press. DOI: https://doi.org/10.1093/acrefore/9780190228613.013.132

Potter, J. (2021). Discursive psychology and naturally occurring talk. In D. Silverman (Ed.), *Qualitative research* (5th ed., pp. 203–222). Sage.

Potter, J. (2002). Two kinds of natural. *Discourse Studies*, *4*, 539-542. https://doi.org/10.1177/146144560200400409

Potter, J. (1996). *Representing reality: Discourse, rhetoric and social construction.* Sage Publications, Inc. https://doi.org/10.4135/9781446222119

Potter, J., & Hepburn, A. (2005). Qualitative interviews in psychology: Problems and possibilities. *Qualitative Research in Psychology*, *2*(4), 281-307. https://doi.org/10.1191/1478088705qp045oa

Potter, J., & Wetherell, M. (1987). *Discourse and social psychology: Beyond attitudes and behaviour.* Sage Publications, Inc.

Potter, J., Wetherell, M., & Chitty, A. (1991). Quantification rhetoric—cancer on television. *Discourse & Society*, *2*(3), 333-365.

Pownall, M. (2025). Bridging qualitative methods and open research. *Nature Reviews Psychology*, 4, 556–557. https://doi.org/10.1038/s44159-025-00477-3

Raymond, G. (2003). Grammar and social organization: Yes/no interrogatives and the structure of responding. *American Sociological Review*, *68*(6), 939-967. https://doi.org/10.2307/1519752

Richardson, E., Stokoe, E., & Antaki, C. (2019). Establishing intellectually impaired victims' understanding about 'truth' and 'lies': Police interview guidance and practice in cases of sexual assault. *Applied Linguistics*, *40*(5), 773-792. https://doi.org/10.1093/applin/amy023

Robinson, J. D. (2007). The role of numbers and statistics within conversation analysis. *Communication Methods and Measures*, *1*(1), 65–75. https://doi.org/10.1080/19312450709336663

Robinson, J. D., Clift, R., Kendrick, K. H., & Raymond, C. W. (Eds.). (2024). *The Cambridge handbook of methods in conversation analysis*. Cambridge University Press.

Rowley, E. (2007). On doing 'being ordinary': Women's accounts of BRCA testing and maternal responsibility. *New Genetics and Society*, *26*(3), 241-250. https://doi.org/10.1080/14636770701701774

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. https://doi.org/10.1080/08351813.2017.1262050

Sacks, H. (1984). On doing "being ordinary". In J. Maxwell Atkinson & J. Heritage (eds.), *Structures of social action: Studies in conversation analysis* (pp. 413–429). Cambridge University Press.

Schegloff, E.A. (1997). Narrative analysis: Thirty years later. *Journal of Narrative and Life History*, *7* (1-4), 97-106. https://doi.org/10.1075/jnlh.7.11nar

Schegloff, E. A. (1993). Reflections on quantification in the study of conversation. *Research on Language and Social Interaction*, *26*(1), 99-128. https://doi.org/10.1207/s15327973rlsi2601_5

Schegloff, E. A. (1992). Introduction. In G. Jefferson (Ed.), *Lectures on conversation* (pp. ix-lxii). Blackwell.

Schegloff, E. A. (1968). Sequencing in conversational openings. *American Anthropologist*, *70*, 1075-1095. https://www.jstor.org/stable/669510

Sikveland, R., Stokoe, E., & Symonds, J. (2016). Patient burden during appointment-making telephone calls to GP practices. *Patient Education and Counseling*, *99*(8), 1310-1318. https://doi.org/10.1016/j.pec.2016.03.025

Speer, S. A., & Stokoe, E. (2014). Ethics in action: Consent-gaining interactions and implications for research practice. *British Journal of Social Psychology*, *53*(1), 54-73. https://doi.org/10.1111/bjso.12009

Stevens, M., Wehrens, R., & De Bont, A. (2018). Conceptualizations of Big Data and their epistemological claims in healthcare: A discourse analysis. *Big Data & Society*, *5*(2), 1-21. https://doi.org/10.1177/2053951718816727

Stivers, T. (2015). Coding social interaction: A heretical approach in conversation analysis? *Research on Language and Social Interaction*, *48*(1), 1-19. https://doi.org/10.1080/08351813.2015.993837

Stivers, T. (2008). Stance, alignment, and affiliation during storytelling: When nodding is a token of affiliation. *Research on Language and Social Interaction*, *41*(1), 31-57. https://doi.org/10.1080/08351810701691123

Stokoe, E. (2021). The sense of a conversational ending. *Nature*, *593* (7859), 347-349. https://doi.org/10.1038/d41586-021-01074-z

Stokoe, E. (2020). Psychological matters in institutional interaction: Insights and interventions from discursive psychology and conversation analysis. *Qualitative Psychology*, *7*(3), 331. https://doi.org/10.1037/qup0000162

Stokoe, E. (2018). *Talk: The science of conversation*. Little, Brown.

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. https://doi.org/10.1080/08351813.2013.780341

Stokoe, E. (2010). "Have you been married, or…?": Eliciting and accounting for relationship histories in speed-dating interaction. *Research on Language and Social Interaction*, *43*(3), 260-282. https://doi.org/10.1080/08351813.2010.497988

Stokoe, E., & Albert, S. (2024). Just a method in search of a problem? The power of conversation analysis. In H. Zhang Waring & N. Tadic (Eds.), *Critical conversation analysis: Inequality and injustice in talk-in-interaction* (pp. 197-223). Multilingual Matters.

Stokoe, E., Antaki, C., Bracher, M., Chrisostomou, L., Henderson, E., Jones, D., & Stewart, S. (2021). The softness of hard data. *Medium*. https://elizabeth-stokoe.medium.com/the-softness-of-hard-data-475743d8a2f2

Stokoe, E., & Edwards, D. (2008). "Did you have permission to smash your neighbour's door?" Silly questions and their answers in police-suspect interrogations. *Discourse Studies*, *10*(1), 89-111. https://doi.org/10.1177/1461445607085592

Stokoe, E., Edwards, D., & Edwards, H. (2016). "No comment" responses to questions in police investigative interviews. In S. Ehrlich, D. Eades & J. Ainsworth (Eds.), *Discursive constructions of consent in the legal process* (pp. 289-317). Oxford University Press.

Stokoe, E., Raymond, G., & Whitehead, K. A. (2025). Categories in social interaction: Unlocking the resources of conversation analysis and membership categorization for

psychological science. *Annual Review of Psychology*, *76*, 531-557. https://doi.org/10.1146/annurev-psych-020124-023147

Storer, N. W. (1967). The hard sciences and the soft: Some sociological observations. *Bulletin of the Medical Library Association*, *55*(1), 75-84. https://pmc.ncbi.nlm.nih.gov/articles/PMC198502/

Tovares, A. V. (2019). On doing 'being ordinary': Everyday acts of speakers' rights in polylingual families in Ukraine. In J. Lee & S. Dovchin (Eds.)., *Translinguistics* (pp. 228-240). Routledge.

Turowetz, J. (2015). The interactional production of a clinical fact in a case of autism. *Qualitative Sociology*, *38*, 57-78. https://doi.org/10.1007/s11133-014-9294-8

Wetherell, M. (1998). Positioning and interpretative repertoires: Conversation analysis and post-structuralism in dialogue. *Discourse & Society*, *9*(3), 387-412. https://doi.org/10.1177/095792659800900300

Waring, H.Z., & Tadic, N. (Eds.) *Critical conversation analysis: Inequality and injustice in talk-in-interaction*. Multilingual Matters.

Whitehead, K.A., Stokoe, E., & Raymond, G. (2025). *Categories in social interaction*. Routledge.

Wieder, D. (1988). From resource to topic: Some aims of conversation analysis. *Annals of the International Communication Association*, *11*(1), 444-454. https://doi.org/10.1080/23808985.1988.11678701

Wilkinson, S. (2011). Constructing ethnicity statistics in talk-in-interaction: Producing the 'White European'. *Discourse & Society*, *22*(3), 343-361. https://doi.org/10.1177/0957926510395446

Wooffitt R. (2007). Communication and laboratory performance in parapsychology experiments: demand characteristics and the social organization of interaction. *British Journal of Social Psychology*, *46*(3), 477–498. https://doi.org/10.1348/014466606X152667

Wooffitt, R. (1991). I was just doing X ... when Y': Some inferential properties of a device in accounts of paranormal experiences. *Text*, *11*, 267-288. https://doi.org/10.1515/text.1.1991.11.2.267