Chapter 4

Natural and contrived data

Simon Goodman and Susan Speer


In this chapter we assess the impact of Jonathan Potter’s (2004) classic text ‘*Discourse Analysis as a way of analysing naturally occurring talk*’. This text has had a major impact on discursive psychology and other qualitative approaches by stimulating methodological debate about the merits and limitations of different ‘types’ of data. It is important to note that the 2004 version of this text is actually an updated version of an earlier piece (Potter 1997), and includes a discussion of the debate the original piece stimulated (e.g. Speer 2002a). Potter’s text itself builds upon his own previous work (Potter and Wetherell 1995) and develops the work of others, notably conversation analysts Sacks (1984) and
Heritage (1984), who had already called for an analytical focus on non-contrived ‘actual events’ (Sacks 1984: 26).

This chapter begins with an outline of the main argument developed by Potter which is that discursive psychologists should focus on data that exists independently of researchers (naturally occurring talk), rather than data that is ‘got up’ (Potter 2004: 205) for the purpose of analysis. Next, we consider how this argument has been developed in subsequent work in discursive psychology and conversation analysis. We argue that the analysis of naturally occurring data can generate fascinating insights into the action orientation of everyday talk. However, developing ideas from our own work, we also suggest that researcher generated and naturally occurring data need not be viewed as discrete ‘types’ in the way that Potter suggests. Like us, Potter acknowledges that ‘contrived’ (2005: 206) data can be ‘naturalised’ (e.g. 2002) and treated as a topic in its own right. However we suggest that this does not go far enough and that all data can be viewed as either natural or contrived depending on what the researcher wants to do with it and what consequences the interactional setting has for the phenomenon being analysed. We conclude by considering the implications of this debate for the ‘natural/contrived’ distinction and future interaction research.

**The argument for naturally occurring data**

Potter defines naturally occurring data as ‘spoken language produced entirely independently of the actions of the researcher’ (2004: 205) and cites a range of examples of types of data that fit this definition: ‘everyday conversations over the telephone, the records of a company board meeting, or the interaction
between doctor and patient in a surgery’ (2004: 205). Naturally occurring talk is contrasted favourably with its alternative, data that is “got up’ by the researcher’ (2004: 205) using traditional research methodological techniques such as interviews and focus groups. Potter, however, explicitly distinguishes naturally occurring interviews, such as broadcast news interviews, from researcher generated interviews applying what he later referred to as the ‘(conceptual) dead social scientist’s test – would the data be the same, or be there at all, if the researcher got run over on the way to work?’ (2002: 541).

Potter (2004) argues that naturally occurring data is preferable to data ‘got up’ by researchers because it is completely free of any researcher influence. Potter deems researcher influence to be problematic for three key reasons: First, researcher influence means the data is ‘contrived’ (2004: 206). Potter regards contrived data as artificial and as something that does not reflect the world as it would be without researcher intervention. Second ‘it is subject to powerful expectations about social science research fielded by participants’ (2004: 206), which suggests a problem with demand characteristics where research participants may act in a particular way because they are aware that they are taking part in social science research. Finally, Potter suggests that ‘there are particular difficulties in extrapolating from interview talk to activities in other settings’ (2004: 206) because ‘contrived’ data are ‘affected by the formulations and assumptions of the researcher’ (2004: 206). Potter’s view reflects a broader concern that ‘researcher effects’ (where the mere presence of a researcher can influence research findings, for example by inducing behaviour that is deemed by the participant to be what the researcher expects, are problematic and to be avoided e.g. Bryman 2004, Coolican 2009).
Potter illustrates his argument for naturally occurring data with examples of naturally occurring broadcast interviews between Princess Diana and the UK journalist, Martin Bashir1 and Salman Rushdie and David Frost, an article from the British Guardian newspaper and a relationship counselling session. He uses this data to demonstrate features of the action orientation of talk in natural settings. Potter focuses on the phrase “I dunno”, showing how it functions as a ‘stake inoculation’ (2004: 212) to fend off suggestions that the speaker has a personal interest (or ‘stake’, see Edwards and Potter 1992) in the things they are saying (Potter 1996), which could potentially be used to undermine their claim.

In sum, Potter’s argument is that research is most effective when being used to address a phenomenon in its natural setting, free from any intervention by a researcher. His argument represents an implicit criticism of the large body of discourse analytic work that has generated ‘contrived’ data.

In later writing Potter and Hepburn (2005, see also Potter and Hepburn 2012) develop Potter’s (2004) argument. Potter and Hepburn's (2005) text is discussed by Rapley (this volume), which demonstrates the wide and ongoing impact of Potter’s (2004) text. Potter and Hepburn further highlight potential problems with researcher generated interviews that can limit their usefulness. They are particularly critical of what they refer to as the ‘failure to consider interviews as interaction’ (2005: 290) whereby the interactional element of the interview is ignored in favour of treating participant responses in isolation from the context of the interview itself. While other researchers may treat interviews as interaction (e.g. Holstein and Gubrium 1995) Potter and Hepburn suggest that

---

1 A verbatim transcript of the interview can be found here
http://www.bbc.co.uk/news/special/politics97/diana/panorama.html
this interaction is rarely properly addressed. Potter and Hepburn consider this lack of focus on interaction problematic, they characterise this limitation as potentially avoidable. They go further, however, by suggesting other unavoidable problems with interviews that have major implications for the use of interviews including that the researchers’ agendas and concepts will always be apparent (or ‘flood’) the interview and that issues of stake and interest - features of talk that have been shown to be central to interaction (and therefore discursive psychology, Edwards and Potter 1992) are rarely addressed fully, which can result in missing what the talk is being designed to do.

Potter and Hepburn (2005) therefore advocate the value of reflexivity as one key element of the discursive approach (Edwards and Potter 1992). From this perspective researchers need to attend actively and reflexively to their own involvement with and creation of data. Potter and Hepburn claim this is a necessary step to conducting systematic analysis of interview data. They present contrived interview data as problematic when there is not sufficient reflexive orientation to how knowledge is produced in interviews. In discursive psychology this reflexive orientation is addressed in the conversational, turn-by-turn, analysis of the ways in which talk is used to perform actions that deal with speakers’ interest and accountability in specific contexts. This sets discursive analysis apart from what other qualitative approaches advocate (cf. Smith, Hollway and Mishler (2005) and Rapley (this volume) for responses to the Potter and Hepburn article).

**Developing Potter’s argument for naturally occurring data**
There is now a range of illuminating discursive findings using naturally occurring data (e.g. Auburn 2005; Benneworth 2009; Edwards and Stokoe 2011; Stokoe 2009). In previous work (Goodman and Speer 2007) we analysed a range of sources of naturally occurring data including television news and debate programmes, newspaper articles, election campaign material and a political speech. By using naturally occurring data, we were able to demonstrate how the classification of asylum seekers was not just something that we were interested in as researchers, but was something that could also be demonstrated to be a concern for the participants. Consider extract 1, below – a statement made by Peter Hitchens, a British newspaper columnist and opponent of asylum seeking during a televised debate programme:

(1) [Goodman and Speer 2007: 170. Asylum: Face the Nation, BBC1, July 23, 2003]

1. the most ... inflammatory language which is used is ... the false
2. use of ... the words ‘asylum seekers’ to describe people who
3. are in fact illegal immigrants

Through analysis of this extract we were able to claim that the categorisation of asylum seekers is something that participants in the asylum debate topicalise themselves. This, in turn, gave us the legitimacy to explore the uses of categories around asylum seeking by participants in the asylum debate without imposing our own definitions and ideas on the data. The analysis of category use showed that references to ‘illegal immigrants’, oriented to the practical action of undermining arguments for asylum seekers, so that this category was most commonly used in opposition to asylum seeker rights.
This example demonstrates the usefulness of focussing on naturally occurring data rather than interview data in which we seek politicians’ views on asylum seeking, for example. By using naturally occurring, instead of contrived, data we were able to identify important elements of a debate without influencing that debate. However, despite the benefits of focussing on naturally occurring data in this way, Potter’s (2004) clear favouring of naturally occurring data and Potter and Hepburn’s (2005) critique of social scientific interviews have proved to be controversial. It is the debates around these arguments that we concentrate on next.

**Challenging Potter’s natural/contrived distinction**

Speer (2002a 516): points to a contradiction between the two positions that state (1) that bias in the form of context effects is not a problem but is a fundamental aspect of all interaction, that can be celebrated and explored and (2) that naturally occurring data is superior to contrived data because it is free of the interviewer's bias/influence.

Speer claims that since all data must be recorded and consent must be sought from all participants to meet ethical standards, it is impossible for any research to be truly free of the researcher. In this sense ‘all data are researcher-prompted and thus contrived’ (2002a: 516, emphasis in original). Speer (2002a: 518) concludes by stating that the distinction between natural and contrived data is ‘inherently problematic’ but that this ‘need not necessarily involve abandoning the concept of ’naturally occurring data’ (which some might argue is still rhetorically useful). Rather, we need to be clearer and more consistent about what exactly constitutes the object of our analyses’ (2002a: 521). Speer’s
(2002a) paper led to a number of responses (Lynch 2002; Potter 2002; Ten Have 2002) and a rejoinder. (We return to this point later when Speer’s proposed solution is discussed.)

In a related challenge to Potter’s argument for a natural/contrived distinction, Griffin (2007a; 2007b) suggests that the concept of natural data is problematic as ‘No talk or other practice is ‘natural’ in the sense of being unmediated by the context of the occasion in which it is generated’ (2007a: 248). Griffin takes issue with Potter and Hepburn’s (2005) suggestion that the research interview is too ‘flooded’ with the researcher’s agenda and instead claims that participants are far more active than this position would suggest, stating this ‘perspective gives the researcher and the research project overwhelming dominance over the research encounter, relegating other participants to relatively passive ‘feeder’ roles’ (2007a: 250). Griffin states that for her the value of researcher interviews is that they constitute a meeting of the researcher's and, importantly, the participants’ agendas. For Griffin this means that analyses should address ‘those points at which the agendas and perspectives of interviewers and interviewees interact’ (2007a: 261) which means that ‘natural’ data totally separate from the researcher is not something to strive for.

We have now outlined two of the main challenges to Potter’s argument for the primacy of ‘naturally occurring’ data. As we demonstrate above, we have used what has been termed ‘naturally occurring’ data in our research. However, alongside Potter (e.g. Wetherell and Potter 1992), we have also used ‘contrived’ data (that is, data that would fail Potter’s ‘dead social scientist’s test) and have actively engaged with our participants. For us, depending on the ways in which
the data is being used, 'naturally occurring' and 'contrived' data need not be viewed as discrete types.

**Naturalising 'contrived' interview data**

We will now demonstrate the ways in which the natural/contrived distinction can be viewed as problematic, with examples from our own analyses, showing how 'contrived' data can be treated as 'natural' depending on the purpose of the analysis. This means that 'contrived' data can be 'naturalised' so that the features of the research setting can become the topic of the analysis and therefore can be treated as 'natural'. Speer (2012; Speer and Stokoe 2014) has demonstrated that contrived materials including social science interviews and focus groups can be 'naturalised', or treated as natural, in ways that contribute to our understanding of interaction (also see Maynard, Houtkoop-Steenstra, Schaeffer and van der Zouwen 2002). It is worth noting that here that we and Potter (2002) are in agreement over the possibility of naturalising data.

Speer and Stokoe (2014) highlight the advantages of naturalising 'contrived data'. The data show how consent gaining (an important part of ethical data collection) can operate very differently in practice from how it is described in theory. The following extracts from the analysis (Speer and Stokoe 2014: 62-63) occurred at the start of telephone interviews with gender identity clinic patients:

(2) [Speer and Stokoe 2014: 62-63 Pre-recorded Message 5. 2:44]

1 Int: Uh::m, I’m happy to go on longer if you are but you
2 must say if you wan- _ need to stop or anything like
3 [that.]
4 Pt: [ Yes.]
5 Int: .hhh Uh:mm pt. t. the interview’ll be _ta:ped if 
that’s okay with [you:
6 Pt: [Yes that’s _fine by me,
7 Int: [.hhh
8 Uh:mm, an’ I’ll write hand-written notes at the same 
ti:me, [.hhh
9 Int: [Uh huh,
10 (3) [Speer and Stokoe 2014: 63. Patient Telephone Interview 2. 00:51]
1 Int: I’m just gonna tick boxes to show that I’ve–
2 I’ve _taken you through this information. 
(0.2)
4 Pt: Okay,
5 Int: Pt. .hhh Uh:mm, pt. so the interview will be ta::ped, with 
7 your con↑se:nt, if that’s oka:y? .hh[h
9 Pt: [Yep.
10 Int: And I’ll write hand-written notes at the same _ti:me,
11 ((Rustling – 0.4))
12 Pt: Mm hmm,

In their analysis Speer and Stokoe (2014: 63) use this seemingly contrived data naturalistically to show how the ‘recording party invites the recipient to confirm that it is ‘okay’ to record (Extract 2, line 5–6; Extract 3, lines 6–7) before continuing delivering information about the study. The recipient subsequently gives their consent (Extract 2, line 7; Extract 3, line 9). They therefore demonstrate the conversational practices of requesting consent constrain opting out of participation.

This next example of naturalisation can be seen in Goodman, Burke, Liebling and Zsada’s (2014) research with asylum seeking refugees. The aim of this project was to understand how refugees talk about their situation in the UK. While issues of asylum and refuge are particularly topical in UK debates (as Goodman and Speer’s (2007) research demonstrates) asylum seekers
themselves tend not to be heard and are generally absent from the discursive literature about this topic. Like Griffin (2007) advocates, Goodman et al. approached their research motivated by their interest in the topic, and specifically what they considered a social/political problem of the mistreatment of asylum seekers. In the following example the interviewer (A3) is asking a participant to elaborate on a point about not wishing to return to her country of origin.


1. A3: So you would never return to Kenya because you would be
2. worried about yourself?
3. P9: =How can go I I face death how can I go I face death? How
4. even if yourself how you can go to a place where you face
5. death (A3: no I know) I can die there it is better I die
6. here better than I go.
7. A3: No you’re right it’s better to be safe
8. P9: Because here I don’t have anything good here I don’t have
9. any life here you understand my life what I explained to you
10. I do not have a good life here but I am safe I stay here
11. because (.) for here I have never been happy even one day
12. here (A3: no) I have never been happy one day

In this extract the participant can be seen to be drawing on the notion of safety as a justification for remaining in the UK. This safety is presented as more important than her happiness so P9 states that she is willing to stay in the UK and be unhappy as this is preferable to having no safety outside of the UK. Importantly, this ‘helps to position P9 as a legitimate refugee as she is placing importance on safety over anything else’ (Goodman et al 2014:28). The interview context can be seen to be directly influencing the interaction in this

---

2 See Goodman et al (2014) for discussion of the three exceptions.
extract. For example the interviewee refers to ‘you understand my life what I explained to you’ (line nine) which references the interview setting where the interviewee is expected to explain her ‘story’ as a refugee to the interviewer. In this case the interviewee is using the interview context, and the previous telling of her circumstances as an interactional resource to bolster her claim about requiring safety above everything else.

The key finding of the analysis is that asylum seekers make complaints about their situation in ways that manage their identity as legitimate refugees. These asylum seekers are therefore orienting to the pervasive suggestion that asylum seekers are economic migrants rather than refugees (an argument identified in Goodman and Speer’s (2007) analysis of what Potter would describe as ‘naturally occurring’ data). In the case of this ‘contrived’ data then, the presence of the interviewers/members of host country meant that questions of legitimacy, which are relevant to the on-going wider asylum debate, are present and especially relevant. Indeed the talk of the asylum seekers in this interview data must be understood in relation (and response) to the televised debate data (Goodman and Speer 2007) described above. This speaks to Speer's (2002) argument that what counts as contrived (interviews gotten up to elicit the stories/accounts of asylum seekers) and natural (in this case real asylum seekers interacting with real British citizens) is not absolute.

Potter agrees that data can be ‘naturalised’ in the ways demonstrated above, where we disagree with him is regarding the implications naturalisation has for the natural/contrived distinction, with Speer suggesting that this allows us to transcend the distinction and Potter rejecting this claim. For Potter naturalisation is not enough to challenge (or ‘blunt’ 2002: 541) the
natural/contrived distinction he makes. Instead Potter argues that ‘the possibility of studying how a particular bit of social research is contrived does not show that the contrived/natural distinction is not useful’ (2002: 541) and that ‘treating method as topic is not the same as using it to find something out’ about the world/our participants (2002: 539). This means that for Potter naturalisation is a beneficial tool, but he rejects Speer’s claim that this renders the natural/contrived distinction problematic, claiming instead that naturalisation is not the same as being able to do effective research in its own right on ‘contrived data’. For Potter, naturalising ‘contrived’ data serves to highlight the influence of these contrived contexts on the interaction, which reflects his critique of ‘contrived’ methods; for him the contrived/natural distinction remains.

**Procedural consequentiality**

Speer further challenges the contrived/natural distinction by arguing that ‘What are natural data and what are not is not decidable on the basis of their type and/or the role of the researcher within the data. All data can be natural or contrived depending on what one wants to do with them’ (Speer 2008: 307, emphasis in original). Speer (e.g. 2002a; 2002b; 2008) argues that rather than the researchers deciding what counts as natural or contrived data, it may make more sense to assess how the participants in a given conversation orient (or not) to the context and to the influence of the researcher). It is argued that decisions about what data to use should be informed instead by the notion of procedural consequentiality, which refers to the impact of the interactional context on the phenomenon of interest. Schegloff says that two questions need to be asked to
help us determine whether a setting is procedurally consequential for a particular phenomenon:

How does the fact that the talk is being conducted in some setting (e.g. ‘the hospital’) issue in any consequence for the shape, form, trajectory, content or character of the interaction that the parties conduct? And what is the mechanism by which the context-so-understood had determinate consequences for the talk. (Schegloff 1992: 111).

To illustrate this, Schegloff (1991) refers to two studies (Levelt 1983 on self-repair and West and Zimmerman 1983 on interruptions) and shows how in the first, but not second, the setting placed limitations on who could talk. The former, as well as the latter, study took place in an experimental setting. Schegloff concluded that the experimental setting in the former study was procedurally consequential for self-repair, while in the latter it was not (Speer 2002a: 520).

This means that decisions about what data may be used for analysis need to be made with consideration given to what influence the context may have on the phenomenon of interest and how talk plays out. In some ‘contrived’ situations, therefore, the contrived nature of the interview may well impact on the phenomenon of interest, whereas in other ‘contrived’ situations it may not. To demonstrate the importance of procedural consequentiality, Speer (2012) conducted a comparative analysis of hypothetical questions in data from four different settings including those that Potter would describe as contrived: ordinary conversations, research interactions, broadcast news interviews, and doctor-patient consultations. In the following extract Speer, as the interviewer, can be seen asking her interviewee a hypothetical question:
(4) [Speer 2012: 358-359. One-to-one, semi-structured interview: 19.3.97: 19
Side A]
1 Int: >So do you think-< marriage would change your leisure in
2 any way.
3 (0.8)
4 Mag: ↑Uhmm:: (2.4) ↑I don’t know I probably wouldn’t go- like I
5 probably wouldn’t go: clubbing as much or (1.2) things like
6 that.
7 (0.4)
8 Mag: But yeah I mean you’d- [>probably coz you’d<] have=
9 Int: [ W h y n o t ? ]
10 (0.8)
11 Mag: hhh Oh just coz I’d be older an’ I’d be- you know you »yo-
12 could think oh you’ve got to be more kind of grown up about
13 things«.
14 Int: But if [they] created like night clubs fo::r (0.8) thirty=
15 Mag: [( )]
16 Int: =year old women. and men. to go to. Or i- would- would that
17 be good?
18 (1.2)
19 Mag: Mm::: as long as y- (well it yo-) »you can just see those
20 kind of places as being full of lecherous old men though
21 can’t you.< You know men who are (. ) divorced or looking
22 for a fling, looking for some: (. ) somebody their of their
23 age whose gonna (1.0) you know (0.4)»it has that kind of
24 i(h)mage about it«. (Sarah goes to) these over twenty-fives
25 things (0.4) »n it’s
26 Int: [Mm,
27 Mag: [full of old people »isn’t it really« <.
28 (0.4)
29 Int: Mm{:}
30 Mag: [I don’t know I think I’d rather just not go after a
31 certain (. ) I’d rather (0.2) say >Right when I hit (0.2)
32 thirty .hh »I’m never gonna go to a club again<’

Speer (2012: 359) shows how her hypothetical question (lines 14-17) is an
‘attempt by the interviewer to probe and unpack the reasoning behind Maggie’s
views on this topic’, which appears to be a generic function of hypothetical
questions across the data. However, Speer also shows that the hypothetical question worked differently in this contrived setting. Here it was found to be ‘more provocative and challenging in its design’ than in other settings (2012: 359). Rather than taking this difference to mean that the interview data should be viewed as ‘unnatural’, this analysis demonstrates that apparently ‘contrived’ interviews are not unique in affecting the way the phenomenon functions. Instead, each setting has its own interactional affordances which are procedurally consequential for the way hypothetical questions run off in each case. This supports our call to determine the use of data based on procedural consequentiality rather than a problematic contrived/natural distinction.

**What is the future of the debate about natural/contrived data?**

The methodological issues addressed here are not only relevant for discursive and conversation analysts, but linguists (e.g. Mann 2011) and qualitative researchers more generally (e.g. Roulston 2010) are starting to question the use of interviews for the reasons Potter has identified. It is worth noting that no parties in these debates have argued that either ‘naturally occurring’ or ‘contrived’ data cannot yield helpful findings. There is (limited) agreement that both types of data can and should be used for analysis. It also seems clear that this debate is not about to be settled, with passionate disagreement on both sides. Our view, developed here, is that a natural/contrived distinction is problematic, because these two types of data need not be viewed as discrete types. What we hope to have demonstrated is that like conversation analysts before him, Potter (2004) developed a strong argument for the use of naturally
occurring data that is supported by the (growing) body of discursive work that has used this data to better understand the way that talk is oriented to practical action in natural settings. Nevertheless, discursive and conversation analytic work continues to use so called contrived data.

We therefore argue that both ‘natural’ and ‘contrived’ data can be used for analysis because, as demonstrated throughout this chapter, analyses of both can yield useful findings. Where we disagree with Potter is that for us this contrived/natural distinction can be transcended and rather than making decisions based on what is ‘natural’ and what is ‘contrived’, we argue that decisions about what data can be used for analysis can be based instead on what the researcher is trying to do with the data and what impact the interactional context has on the phenomenon being studied, focussing on the procedural consequentiality of the research context. Potter’s call for analysis of naturally occurring data, and the novel findings that it has brought to discursive psychology, can only be a good thing for the discipline and we fully expect a focus on naturally occurring data to increase over the coming years. Alongside this, interviews continue to be used and, as Potter (2007) has argued, will most likely continue to dominate.

The debate over which kind of data is best, or as we’ve argued here, if it is useful to talk in terms of two distinct types of data at all, will continue. This is likely to encourage a more nuanced understanding of the ways that methods and the context of data relate (Speer 2002: 521) and it may also encourage more social scientists to analyse interview data as an interaction. This is important as it can guide researchers to make methodological decisions about appropriate data for their research questions and to consider the impact of the context in
which data is produced. It is likely that this debate will reignite again, if or when it does there is scope for further development into exactly what impact the interview context and the social science interviewer has on the ways participants manage their stake and interest in interviews. Any further debate has the possibility of developing and increasing methodological rigour.

References


