

Enhancing researcher reflexivity about the influence of leading questions in interviews

Introduction

Interviewing remains the most common qualitative method of data collection, reflected in a substantial literature including numerous handbooks on interviewing (Alvesson, 2003; Gubrium & Holstein, 2011) as well as critical appraisals (Langley & Meziani, 2020). Around a third (78) of the research articles published in this journal between 2011 and 2021 have used interviewing; for example, sensemaking (Chreim & Tafaghod, 2012), metaphors of change agency (Cassell & Lee, 2012), identity (Foldy, 2012; Jacobs, Oliver, & Heracleous, 2013), different facets of change (Ford, Lauricella, Van Fossen, & Riley, 2020) and change leaders behavior (Higgs & Rowland, 2011). Furthermore, interviews are not only confined to studies that use them for data collection, they are the basis for theorizing for example Schein's sociopsychological model about change (Coghlan, 2021) and they are also used to understand and scope the research context as a precursor to studies that use surveys (Coruzzi, 2020).

Despite the apparent simplicity of employing everyday skills of language and conversation, interviews are "very hard to do well" (Potter & Hepburn, 2005, p. 300). They are complex social situations that involve meaning-making, emotional and identity work, impression management, political action or the application of cultural scripts (Alvesson, 2003). Interviewers are "an essential part of the trajectory of the talk" (Rapley, 2012, p. 541) who have considerable power (Kvale, 2007; Tanggaard, 2007) because they frame interviews; they ask the vast majority of questions and they decide what constitutes relevant and sufficient responses (Wang, 2006).

This paper is concerned specifically with ways in which interviewers can unduly influence, and therefore bias, interviewees' responses through the wording of interview questions. While the general problem is acknowledged by authors including Mishler (2009), Patton (2015) and Langley and Meziani (2020), detailed analysis of how the problem arises in practice is lacking. Attention to the way questions are formulated is vital because "the form of the answer may be occasioned by the form of the question" (Potter & Hepburn, 2005, p. 588) especially as respondents typically do not

notice the suggestibility of a question or the effect that specific words may have on them (Hubbard, 1950).

The purpose of this paper is to expand understanding of how interview questions can lead or bias responses and cast doubt on the trustworthiness of findings. It presents a new typology of ways in which the wording of interview questions can lead an interviewee's response. The typology informs a method, the 'cleanness rating', through which researchers can address and mitigate this problem, improve their questions and enhance researcher reflexivity. While we mostly refer to questions including those in pre-designed protocols, variations to the protocol and questions spontaneously generated in interviews, this paper relates to the wording of interviewers' verbal interventions including statements.

Despite the acknowledged importance of the influence of questions on interviewees' responses, the field lacks conceptualization of how, specifically, the wording of questions can do this. Instead, the literature tends to rely on emphasizing "the need to avoid leading questions" (King, Horrocks, & Brooks, 2019, p. 54). Little specificity exists beyond Patton's assertion that "leading questions ... give the interviewee hints about what would be a desirable or appropriate kind of answer" (2015 Loc 15075).

The consequences of this influence are important for the trustworthiness (Lincoln & Guba, 1985) of interview-based research, specifically their confirmability which:

"... is concerned with providing evidence that the researcher's interpretations of participants' constructions are rooted in the participants' constructions and also that data analysis and the resulting findings and conclusions can be verified as reflective of and grounded in the participants' perceptions. In essence, confirmability can be expressed as the degree to which the results of the study are based on the research purpose and not altered due to researcher bias." (Jensen, 2008, p. 112).

Confirmability therefore concerns the potential for the constructs and meanings presented in research findings to be derived more from the interviewer's questions than the interviewee's responses. In other words, how do we know that the interviewee is the 'author' of the data from which the interviewer derives their findings? This affects confidence in findings that claim to

authentically represent participants' meanings. In this paper we refer to this as an issue of *authorship*, in that the wording of questions can cast doubt on whether the interviewee or the interviewer is the 'author' of the data. Additionally, the notion of authorship draws attention to the point at which data are collected, whereas a review of confirmability might be postponed until end of the research process.

A focus on the authorship of data enables researchers to develop a more reflexive approach to interviewing (Schaefer & Alvesson, 2020). Thus ideally a researcher should "reflect on how he or she has influenced the situation and the people being studied ... so as to minimize any distorting effect on the research findings." (Hammersley, 2004, p. 934). There are many forms of reflexivity (Cunliffe, 2003; Johnson & Duberley, 2003; 2015), our focus is on the interviewer's awareness of the wording of questions and their potential effect on interviewee responses.

The existence and extent of the problem of the influence of interview questions on data are obscured by the fact that interview transcripts which show the actual questions asked and the interviewee's responses seldom appear in journal articles. Even authors who publish their interview schedules (Isabella, 1990; Petriglieri, Ashford, & Wrzesniewski, 2019; Rogers, Corley, & Ashforth, 2017; Sonenshein, 2010) acknowledge that variations on protocol questions may be used in practice yet these are rarely reported. This makes it hard to judge whether, and to what extent, an interviewer's questions may have biased their data.

Bluhm, Harman, Lee and Mitchell (2011) found, in a review of 198 qualitative articles from the US and Europe in the field of management, that 84% used interviews for data collection and 45% of all articles were "incomplete, missing or extremely vague" in descriptions of data collection and analysis (p. 1880). This lack of methodological transparency (Aguinis, Ramani, & Alabduljader, 2018) has prompted Schaefer and Alvesson (2020) to call for researchers to review how data are collected in order to be clear about the influence of the interviewer on the data. It should be noted that a range of perspectives exist towards methodological transparency, from more positivist approaches that stress the need for replicability (Aguinis & Solarino, 2019) to those which emphasize that transparency signals trustworthiness (Pratt, Kaplan, & Whittington, 2020).

Before proceeding further it must be acknowledged that interviews may be based on a wide variety of ontological and epistemological assumptions (Alvesson, 2003; Langley & Meziani, 2020). Those stances vary in the ways in which interview data are produced (Hammersley, 2017), notions of quality (Roulston, 2010) and the type of knowledge claims likely to be made (Schaefer & Alvesson, 2020). In some conceptions of interviewing, data are regarded as co-authored or co-constructed, hence the influence of questions may be less important.

Authorship is particularly relevant for research whose findings are claimed to represent interviewees' subjective frames of reference and world views, unmediated by the interviewer's constructs. Gill (2014) highlights the potential of a phenomenological approach for the field of organizational change in order to investigate, for example, "institutional work and organizational identity" (p. 118). While phenomenological methodology has several variations (Gill, 2014), interviewers typically guide introspection (Vermersch, 2009) in order to investigate a person's lifeworld and subjective meaning (Eberle, 2014). Researchers are encouraged to practice 'epoché' (Husserl, 1982; Moustakas, 1994), which refers to setting aside or bracketing their own assumptions about a phenomenon to enable the interviewee to describe the phenomenon in their own words. Conklin (2014) acknowledges the difficulty for phenomenological researchers of practising epoché, noting that the researcher's questions inevitably create structures and categories for the participant; yet few guidelines exist for how to guide introspection while also bracketing assumptions (LeVasseur, 2003).

In summary, the potential for the wording of interview questions to lead responses can cast doubt on the authorship of data and the trustworthiness of findings. This problem is important, given the prevalence of interview-based research in organizational change, yet is under-conceptualized and compounded by a lack of methodological transparency in that interview transcripts are seldom made available for scrutiny in published research.

The article proceeds as follows. It begins with a typology of three features of questions that can lead responses. We then introduce the cleanness rating, a method that enables researchers to become more reflexive about how their questions might lead interviewees' responses. The rating is

Accepted manuscript version of:

Cairns-Lee H, Lawley J, Tosey P. Enhancing Researcher Reflexivity About the Influence of Leading Questions in Interviews. *The Journal of Applied Behavioral Science*. August 2021. doi:10.1177/00218863211037446

illustrated with examples from an interview-based study of business leaders' conceptions of leadership before we conclude with a discussion and possible limitations of this approach.

How Questions Lead; A Typology

What constitutes a leading question and how can this be identified? Many authors agree that leading questions are undesirable (Brinkmann & Kvale, 2014; Creswell & Poth, 2018; Rubin & Rubin, 2005). When definitions are attempted they often merely recapitulate the characteristics or effect of such a question; for example, "its wording suggests to the interviewee the kind of responses that is anticipated" (King et al., 2019, p. 51) or "those that are phrased in a way that leads the interviewee in a particular direction" (Yeo, Legard, Keegan, Ward, McNaughton Nicholls, & Lewis, 2014).

We could find no analytical account in the organizational literature that would enable a researcher to distinguish between leading and non-leading questions. In short, there is a plethora of descriptive treatments which seem to cast the issue as one of common sense and emphasize the need to ask clear, open questions instead (e.g. Patton, 2015). Yet designing clear and insightful questions is more difficult than many researchers realize (Qu & Dumay, 2011). As Patton himself says, with reference to a classic book on questioning by Stanley Payne (2015, p. 3), "If it all boils down to the familiar platitudes about using simple, understandable, bias-free, nonirritating wordings, all of us recognize these obvious requirements anyway. Why say more?" (Patton, 2015 loc. 14232).

There is indeed a need to say more, precisely because the influence of question wording on interviewees' responses can be far from obvious and the potential for researchers to use leading questions *unintentionally* is likely to be underestimated. A typology, presented in Table 1 and followed by more detailed description illustrated with examples, specifies three features of questions that have the potential to lead: introduced content, presuppositions and evaluation.

The typology is derived from and supported by theory concerning the role of metaphor (Hatch & Yanow, 2008; Lakoff & Johnson, 1980), theory on presuppositions (Beaver & Geurts, 2014), experiments in how witnesses can be led during an interview (Sharman & Powell, 2012) and theory on the acquiescence effect (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003).

[Table 1 Leading features of interview questions](#)

Introduced Content

Introduced content refers to terms used by the interviewer that have not previously been used by the interviewee. Gioia (2021, p. 23), for example, recounts the problem that arises from “trying to use our concepts to understand others’ ways of understanding”. Introduced content is identifiable directly by the actual words used in a question.

The role of metaphor in interview questions is largely overlooked in the literature on qualitative research. Far from being occasional, poetic figures of speech, metaphors are fundamental to human cognition, ubiquitous in thought and language and central to how people think and talk about things (Gibbs, 2008). According to Lakoff and Johnson, “*The essence of metaphor is understanding and experiencing one kind of thing in terms of another*” (1980, p. 5 italicised in original). In the cognitive linguistic view, humans commonly use a process of metaphorical extension to draw from a source domain of “journeys, war, buildings, food, plants and others” to conceptualize target domains such as “life, arguments, love, theory, ideas, social organizations and others” (Kovecses, 2002 p. 4). Metaphors act as “framing devices” (Hatch & Yanow, 2008, p. 36) through which people make and manage meaning (Smircich & Morgan, 1982). Hence “metaphors matter” (Oswick, Grant, & Oswick, 2020, p. 287). When used in questions, this framing role effectively narrows the scope of an interviewee’s responses.

Metaphors are not only “pervasive in everyday life” (Lakoff & Johnson, 1980, p. 3) they are also often subtle and may go unnoticed. For example, in the question, “How has this memory affected your life? What kind of impact has it had on your life?” (Englander, 2012 p. 31), “impact” is a metaphor. It is a prime example of an experience from the physical world – the action of one object coming forcibly into contact with another – serving as a way to comprehend a more abstract phenomenon – a non-physical marked effect or influence. The metaphor “impact” is just one example of the common “event structure metaphor” (Lakoff, 1993 p. 220) where causes are conceptualized as forces (Johnson, 1987). What makes metaphor especially important for interview questions is that “people are not always aware that they have been influenced by a

metaphor” (Thibodeau, Hendricks, & Boroditsky, 2017, p. 852) because many metaphors are subtle and inconspicuous.

Presupposition

Patton (2015 loc. 14240) cites Payne (1951, p. 16) as saying; “If all the problems of question wording could be traced to a single source, their common origin would probably prove to be in taking too much for granted. ... Frequently, our assumptions are not warranted.” This captures the problem with presuppositions (Beaver & Geurts, 2014), they contain tacit assumptions.

Presuppositions are more subtle than introduced content because they are implicit in the wording of a question. The key to understanding the role of presupposition is to recognize that *in order to make sense of a question the interviewee has to accommodate the presupposition* (initially at least) and they are easily taken for granted by the listener because they pass unnoticed.

Duži and Čihalová (2015) state that the subject of presupposition is complex, there are many ways to categorize them and questions are fraught with presupposition. Indeed, probably all questions involve some presupposition. It is the degree of scope-narrowing or frame-specifying that makes them more or less likely to lead the interviewee. Some are obvious and are acknowledged in the literature as leading questions. Patton (2015) acknowledges presuppositional questions and describes a constructive use for them, which is that they avoid confronting the interviewee with having to make an admission. However, as the examples below demonstrate, in addition to yes/no and either/or questions, more subtle presuppositions exist as potential sources of leading that are rarely identified in the literature.

Due to the taken-for-granted nature of presuppositions they have received little attention in their influence on questions (Duží & Čihalová, 2015) and they may be hard to spot. Here we identify two types that are particularly useful for recognizing presuppositions in interview questions: structural and logical.

In *structural presupposition* the syntax of the question presupposes a situation that the interviewee has not previously stated exists. For example;

What else has changed?

Was it hard to go back to work then?

Do you worry about the future at all?

These questions presuppose, respectively, that something else has changed; that it may have been 'hard' (a metaphor) to go back to work; and that the interviewee might worry about the future (notice that by adding 'at all' to the question the interviewee is invited to find even the slightest worry). In each case the question suggests an answer, and unless the interviewee rejects the premise of the question, the range of possible answers available to them is specified by the presupposition in the question. The problem occurs when the interviewee accommodates the presupposition. Then authorship becomes doubtful since the interviewee may be complying with the presuppositional suggestion in the question, rather than describing their own experience.

In *logical presupposition* the question assumes there is a relationship between two or more items not specified by the interviewee, which is often an assumed cause-effect relationship. For example:

How did that make you feel?

How has this memory affected your life?

Is the relationship important to the outcome?

If the interviewee had not previously stated that they were 'made' to have a feeling as a result of something happening; or that this memory has 'affected' their life; or that an outcome was influenced by the relationship; then (unless the interviewer's presupposition is rejected), the interviewee may well adopt the assumption of the presupposed situation or relationship as part of their answer.

Evaluation

The third way that questions can lead is through the interviewer's introduction of an opinion, an evaluation, or when they raise a doubt or objection about something said by the interviewee. For example:

Really, nothing's changed?

Goodness, how do you manage to connect with so many people?

Exactly right!

These kinds of comments and questions can undermine an interviewee's own opinions or confidence in the value of their experience and make them more susceptible to later leading questions and statements. This may be compounded by two common biases identified in behavioral research; the social desirability effect that "refers to the need for social approval and acceptance" (Crowne & Marlowe, 1964, p. 109) and the acquiescence effect (Podsakoff et al., 2003) which refers to the tendency for respondents to agree with questions. This is not just a concern for inexperienced interviewers, since there is some empirical evidence that "more experienced interviewers obtain higher levels of acquiescent reports than do inexperienced interviewers" (Olson & Bilgen, 2011 p. 99). In addition, the authority that interviewers can be perceived to possess may increase the difficulty for interviewees to reject introduced content, presuppositions or evaluations (Wang, 2006).

Applying the Typology to An Interview Protocol

Note that several of the above examples involve more than one of the three leading features, which can compound the effect, and that several are 'open' questions, as typically recommended in interviewing guidance. The designation of a question as 'open' or 'closed' is independent of its potential to lead an interviewee's answer; contrary to what is often claimed, open questions are not inherently more virtuous than closed questions.

To illustrate the application of the typology we examine an interview protocol published as an addendum to an article in *Administrative Science Quarterly* (Petriglieri et al., 2019). We are grateful to the authors for providing a detailed interview protocol. Most of their questions introduce only the minimum of concepts, with few leading presuppositions. For example (question numbers come from the protocol) :

1. Can you tell me a little bit about what you do?
3. How long have you been working in the current arrangement?
4. How will you know if you are successful in your job?

However, a few of the questions demonstrate one or more of the features from the typology, two examples are:

2. How much freedom do you have in determining the final product of your work?

10. Many people engage in self-talk. What kinds of things do you say to yourself while working?

Using the typology in Table 1 we can see that Question 2 introduces the metaphor “freedom” which may or may not be the way the interviewee evaluates their work, and it presupposes that the amount of freedom “determines” the final product. This is an example of the introduction of a subtle metaphor – freedom – and cause-effect logic that may be difficult to disentangle in the moment.

The preamble to Question 10 makes a factual statement before presupposing that the interviewee is one of the “many people”. If the interviewee happens to be among the more than 20% of people who say they do not engage in self-talk (Goldstein & Kenen, 1988) they might manufacture an answer if they were not prepared to challenge the validity of the question – an action few seem likely to take when being interviewed by a professor.

Without seeing the transcripts of the interviews we cannot know how the 38 interviewees responded to these questions. However, there is a possibility that some of the interviewees will have accepted the metaphors and the presupposed logic of the questions, even if these did not fit their worldview, and hence they could have been inadvertently led. Having been sensitized to the features in the typology, these two questions could be reworded; for example (possible additional questions in italics):

2. Who or what influences the final product of your work? [*How much of an influence do they have?*]

10. Do you have conversations with yourself while you are working? [*If yes, What kinds of things do you say to yourself while working?*]

The Cleanness Rating

The typology identifies and sensitizes researchers to three features of questions that can potentially lead interviewees’ responses. We have illustrated how this can be used to review questions designed for an interview protocol. However, even a well-designed protocol does not guarantee that it will be followed or that its questions will not be altered during the interview. Furthermore, additional questions introduced during the interview may constitute a substantial

proportion of the questions asked. We maintain that unintended leading is especially likely to occur through the generation of spontaneous questions in an interview, since the interviewer has little time to craft such questions. We would also argue that the most effective way for an interviewer to enhance reflexivity about their role in the production of data is likely to be through reviewing the interview questions they actually asked in the light of interviewees' responses. In this section we present a method for this purpose, called the 'cleanness rating' (see Table 2 below).

Whereas the typology (Table 1) identifies the *features* that contribute to leading, the cleanness rating enables an interviewer to assess the *degree* of leading of a question, according not only to the presence (or otherwise) of those three features but also the evidence of the interviewee's response. The rating originates in and has been developed through the method of clean language interviewing introduced by Tosey, Lawley & Meese (2014), which is an application of the principles of clean language (Grove & Panzer, 1989) to interviewing. Originally conceived by New Zealand counselling psychologist, David Grove, clean language is a practice that has been applied subsequently in many non-therapeutic contexts such as coaching (Doyle & McDowall, 2015), education (Groppe-Wegener, 2015) and organizations (Barner, 2008). Grove facilitated his clients to focus on their inner metaphorical world. He devised ways of communicating that he termed 'clean', a metaphor to indicate that his questions minimized the potential for the therapist to 'contaminate' the client's experience. Owen (1996) appears to have been the first to argue that its capacity to minimize the introduction of the interviewer's assumptions makes clean language highly applicable to phenomenological interviewing. Not only is it designed to stay as close as possible to the interviewee's lifeworld, it also provides more specific guidelines than are generally available in the research literature on how to pose questions in order to assist bracketing and achieve epoché.

The cleanness rating method was designed originally to review the extent to which the interviewer adhered to the principles of clean language interviewing (CLI) in a project to elicit six managers' conceptions of their 'work-life balance' (Tosey et al., 2014). It has continued to be developed by being used in this way in studies that intentionally employ clean language as an interviewing discipline (Cairns-Lee, 2017; Linder-Pelz & Lawley, 2015; Nehyba & Svojanovský, 2017).

The cleanness rating enables researchers to assess questions by assigning them to one of six categories, as shown in Table 2, which specifies how each relates to the features identified in the typology (Table 1). The categories effectively form a continuum from questions that are as 'clean' as possible at one end - hence with minimal influence on authorship - to questions that are increasingly more 'leading' and likely to have influenced authorship at the other. Thus 'clean' and 'leading' are not mutually exclusive or dichotomous, rather they are indicative of a relative degree of constraint or freedom for interviewees to respond from their own lifeworld.

The categories are explained below and illustrated by examples taken from Cairns-Lee's study (2017) of 30 business leaders from 15 nationalities that used CLI to elicit their implicit leadership theories (Schyns, Tymon, Kiefer, & Kerschreiter, 2013) and from a paper on interview technique (Englander, 2012).

Table 2: Definition of the six categories in the cleanness rating

Classically Clean

The first category, *classically clean*, comprises a small set of questions developed by Grove and deemed relevant for interviewing by subsequent authors in the field. They have become central to clean language interviewing because they are regarded as being as independent as possible of a practitioner's own terms and presuppositions. This is because they are based on the content of the interviewee in conjunction with 'universal' constructs of time, space and form commonly used to organize experience; in other words, semantic primes – concepts that cannot be expressed in simpler terms and can be regarded as universal in that they appear in every language studied (Goddard & Wierzbicka, 2014). Apart from semantic primes and the interviewee's own words, these questions are devoid of introduced content, presuppositions and evaluations. They are shown in Table 3, where 'X' and 'Y' denote the interviewee's own exact words; for example (Cairns-Lee, 2017):

Interviewee: Leadership, the way I see it, is creating a vision where you want to go to and then lead your company, lead your employees, explain to them where we're heading to, how we're going to do that, what we need for that, what I expect from them to be successful and then giving the example yourself.

Interviewer: And is there anything else about that vision that you create?

This question was rated 'classically clean' because it incorporates only the interviewee's concepts, using their words 'vision' and 'create', in a format that appears in Table 3. This question invites the interviewee to attend further to, discover and describe more about, their model of leadership.

[Table 3: Classically clean questions adapted from Tosey et al., 2014 p. 631](#)

The format of the questions in Table 3 is reduced to their essential form, without context, and therefore may appear stark. Variations in the questions are possible and the delivery of the question is important. Grove (1989) was particularly attentive to the vocal quality including tone, rhythm and pace. Clean questions are best delivered in a tone "of implicit acceptance, curiosity and wonder" (Lawley & Tompkins, 2000, p. 80). The syntax of clean questions is another feature of their formulation and delivery. For example, starting with 'And' conjoins the interviewee's response and indicates that the interviewees' words will be the point of departure for the next question. Repeating some of the interviewee's exact words serves a number of functions: it builds rapport by acknowledging and validating their description and signals that the interviewer is listening acutely.

Clean Repeats

The third category in Table 2, *clean repeats*, accurately recap a selection of the interviewee's content, with precise repetition of the interviewee's words including their use of metaphor. For example:

Interviewee: Leadership the way I see it is creating a vision where you want to go to and then your company, lead your employees, explain to them where we're heading to, how we're going to do that.

Interviewer: Okay. And it's creating a vision about where you want to go to and explaining how to get there.

Interviewee: Yeah. If you talk about that vision, you have your own ideas, but also the vision is created not just by one person or by myself, but just in discussions with the people.

Clean repeats are likely to serve a number of functions. They let the interviewee know that they have been meticulously listened to by acknowledging the interviewee's experience just as they have described it. This repetition helps interviewees to re-hear what they have said, encouraging them to reflect further on their constructs, often dispensing with the need for a question and thereby minimizing potential unintended contamination by the interviewer's implicit assumptions.

It is important to emphasize that clean repeats, like classically clean questions, use an interviewee's exact words. They do not paraphrase because paraphrasing, even if apparently innocuous, is one means by which an interviewer can introduce content that leads subsequent responses. This is one distinction between the principles of clean language and Rogers' (1945) 'nondirective technique for social research', which it could appear to resemble. Rogers claims that this method does "nothing to bias the material... as there have been no questions to guide the interview. There are no evaluations which arouse defence or which shut off expression" (1945, p. 280). An examination of an interview extract shows that despite the intention to "hold up... a verbal mirror" (Rogers, 1945, p. 279) paraphrasing shifts attention from the interviewee to the interpretation of the interviewer-counsellor. For example (Rogers, 1945, p. 280):

MRS. S.: I was just thinking, I could explain all this without coming to a psychologist.

COUNSELLOR: You feel that you really understand the situation.

The counselor's paraphrase transposes Mrs. S's 'was just thinking' to 'feel' and her 'could explain' (which entails doubt) to 'really understand' (which appears to eliminate doubt). This statement introduces the interviewer's own assumptions and appears to imply an evaluation. We would therefore regard it as leading. This is not an isolated example in Rogers' paper. Whilst the intention to enable the interviewee to experience being deeply understood is similar in the non-directive approach and clean language interviewing, the two approaches differ markedly with regard to paraphrasing.

Contextually Clean

There are two types of 'contextually clean' question – topically clean and logically clean. One way in which researchers are almost bound to introduce their own content is in the introduction of the interview topic, therefore sensitivity is needed to avoid superimposing the researcher's

understanding, language or categories. A *topically clean* question aims to introduce a topic which is a prerequisite for the study whilst minimizing the introduction of unnecessary content or leading presuppositions in order to avoid unwanted influence on the authorship of data.

For example, Cairns-Lee (2017) was enquiring into leadership, which is neither a universal construct nor a term originating with the interviewee. The topic was introduced with a contextually clean question; 'What is leadership?' The question is stripped back to its essence and it might appear a little stark in print. However, when asked in a friendly, curious tone, it elicited relevant responses from all 30 interviewees.

A second contextual justification for a question is congruence with the interviewee's logic. In other words, *logically clean* questions are particular to the local context of the interviewee's lifeworld, as distinct from classically clean questions which are generic. An example from the Cairns-Lee study (2017) is:

Interviewee: The first thing that comes is sensing who is this person and what is the situation and what they're trying to achieve. Then I adjust my style of how I talk to them ... It's like watching the sci fi or spy thrillers and somebody walks in and scans the environment ... and then you notice something, all right, this is an important thing. ... essentially, you're just scanning. Tom Cruise 'Mission Impossible' (laughs).

Interviewer: And when you're scanning like Tom Cruise in Mission Impossible, is there something that you're looking for when you're sensing?

Interviewee: Either a connection point or a warning sign.

Here the interviewer introduces the phrase 'something that you're looking for'. Although the interviewee had not specifically said this, they had said they were 'sensing' and 'scanning' like Tom Cruise in the movie. The question was congruent enough with the interviewee's logic for it to be considered contextually clean.

Mildly Leading

Questions that potentially direct the focus of the interview away from the interviewee's lifeworld by introducing explicit or implicit interviewer assumptions are labelled 'leading'. *Mildly*

leading refers to an intervention that, despite a potential bias, seems not to compromise the authorship of the interviewee's answers; for example (Cairns-Lee, 2017):

Interviewee: The way I look at it is that other people saw a certain capability, certain capabilities in me that I was able to do that or something like that.

Interviewer: And did you see those capabilities as well?

Interviewee: No, not then, because for me it was normal.

The question is leading because although it makes use of the interviewee's words 'saw' and 'capabilities', the structure 'did you ... as well' supplies an answer by presupposing that the interviewee *did* see those capabilities. It also shifts attention from how other people saw certain capabilities in the interviewee and introduces the idea that the interviewee may have seen these capabilities in himself. In this case, the interviewee replies in the negative, effectively rejecting the interviewer's assumption. The question appears not to have biased the response; therefore it is categorized as mildly leading.

Strongly Leading

Questions that suggest or imply an answer and substantially narrow the options available to an interviewee are classified as *strongly leading* especially when the interviewee's response provides evidence to doubt the authorship of the answer. They are usually based on interviewer assumptions from outside the logic of the interviewee's descriptions.

The following example is an extract from an academic paper on interview technique (Englander, 2012), which warns generally about 'leading the participant' (p. 26) but makes no mention of the potential for leading in the sample transcript provided:

Interviewer: How has this memory affected your life? What kind of impact has it had on your life?

Interviewee: My dad's girlfriend's apartment or my grandmother? Both?

Interviewer: The first memory. How has this impacted, what impact has it had on your life?

Interviewee: ... I'm not going to attribute it to this memory solely, But it definitely has a very large impact ...

(Englander, 2012 pp. 31-32 n.b. these extracts are separated by authorial comment – not relevant here – following the the first interviewee response, but appear to be contiguous in the interview itself.)

Notice that the interviewer introduces “impact” (a metaphor, as explained above) three times in quick succession. It is highly probable that this is an example of *introduced content* since this question and response are preceded by lengthy extracts from the same interview from which the word “impact” is absent. In addition, the questions demonstrate how both types of *presupposition* can work together to lead an interviewee. In this case, the “How has this ...” and “What ... has it had” structures of the questions combine with a cause-effect logic to presuppose that the memory has ‘affected/impacted’ the interviewee’s life; and this calls into doubt the authorship or provenance of the interviewee’s statement, “it definitely has a very large impact”. This is sufficient to assign both questions to the strong leading category. This example also illustrates that, far from being restricted to novice interviewers, unintended leading can be an issue even for acknowledged experts in qualitative research. ‘Leading’ is by no means confined to rare, exceptional cases of unskillful questioning.

Other

A final, miscellaneous category, *other*, refers to verbal reinforcement words and sounds (mhm, uh-huh, okay) that demonstrate the interviewer is paying attention and encourage the interviewee to continue, as well as neutral responses to questions of clarification from the interviewee; for example:

Interviewer: And what kind of leader would you say you are? (contextually clean)

Interviewee: It depends whether that’s from my own personal view or what do I think people think?

Interviewer: I’ll ask you that in a minute.

Here, the interviewer statement simply gives the interviewee information about the process of the interview. Even so, there can be a fine line between neutrally encouraging an interviewee to continue and subtly implying approval of what they are saying. Words such as ‘right’ and ‘great’ can be used to encourage the interviewee but they could be experienced as evaluations, implying that

the interviewee's answers are being assessed as good/bad or right/wrong. Also, these words may invoke the acquiescence effect (Podsakoff et al., 2003) whereby an interviewee (unconsciously) looks for cues from the interviewer about how to answer.

Applying the Cleanness Rating

To apply the rating, the transcript reviewer considers whether any of the features in Table 1 were present in each of the interviewer's interventions; and if they were, to consider, from the interviewee's perspective and based on their response, how suggestive or leading these interventions might have been. Since we cannot know how an interviewee would have responded if asked another question, we need to examine the interviewee's prior and subsequent responses to consider the evidence for the influence of the question on the authorship of the data. The cleanness rating can be applied to individual questions, to interviews and to whole studies (sets of interviews).

To illustrate its application to a study, Cairns-Lee (2017) rated transcripts of 30 interviews. The results showed that 88% of the interviewer's interventions were allocated to a clean category and 12% to a leading category. Of the leading interventions, 10% were assessed as mildly leading and 2% strongly leading. The proportion of clean to leading cited here is compatible with the proportions obtained in other research using well-trained clean language interviewers (Lawley, 2017; Linder-Pelz & Lawley, 2015; Nehyba & Svojanovský, 2017; Tosey et al., 2014).

We emphasize that these figures relate to studies that have used CLI and are not intended as a general standard. Nevertheless, knowing the relative proportion of leading questions could be a way for researchers to report concisely on confirmability. Based on available transcript extracts examined by the authors (Englander, 2012; Hurlburt & Heavey, 2006; Marshall & Rossman, 2015; O'Broin & Palmer, 2010), a rating of 50% or more leading questions is not uncommon.

Finally, it is worth pointing out that adopting some classically clean questions and clean repeats in the exploration of interviewee responses during an interview can support researchers to avoid leading questions in data collection whilst honoring the interviewee's sense-making. For example Cairns-Lee (2017) found that four questions comprised 38% of all questions asked in her study and clean repeats accounted for a further 29%:

And is there anything else about (...)?

And what kind of (...)?

And how do you know (...)?

And that's (...) like what?

While the fourth question is most suitable when eliciting interviewee-generated metaphors, the first three could be incorporated into many different types of interview to keep attention on the interviewee's experience and invite them to reflect on and elaborate their responses.

Discussion and Limitations

In this discussion we consider the benefits and limitations of the typology of features of leading questions and the cleanness rating. We aim to broaden ways of thinking about researcher reflexivity in interviews and the taken-for-granted nature of questions.

Adopting a reflexive approach has been advocated broadly for researchers (Cassell, Bishop, Symon, Johnson, & Buehring, 2009). We suggest that considering the influence of interviewer questions may be of particular interest to qualitative research about the subjective nature of organizational development and change – the topic focus of JABS – for example, mindset (Maxton & Bushe, 2018), resilience (Gover & Duxbury, 2018), sensemaking (Sharma & Good, 2013), identity (Akram, 2013; Foldy, 2012; Kram, Wasserman, & Yip, 2012), commitment to change (Ford et al., 2020), intrinsic motivation (Auger & Woodman, 2016), cultural competence (Foldy & Buckley, 2017) and psychological contracts (Sverdrup & Schei, 2015).

While the philosophy behind clean language emphasizes the benefits of minimizing inadvertent leading in order to focus on the interviewee's content, it is important to emphasize that bias cannot be eradicated. In clean language it is acknowledged that the interviewer is inevitably directive, for example through the choice of what to attend to and which question to ask. It is part of the skill of the interviewer to direct attention in order to elicit relevant and interesting data. This represents another point of departure from Rogers' (1945) claim to have devised a non-directive method.

There is a fine yet vital distinction to be drawn between staying as close as possible to the interviewee's 'model of the world' and taking a positivistic stance in which it is assumed that the

interviewee's content has an objective reality and is merely waiting to be gathered (Roulston, 2010). The stance of clean language is that a person's lifeworld emerges and evolves through being explored. It is not pre-formed, needing only to be mined. Facilitated by the interviewer's questions, the interviewee elaborates their account through noticing details and relationships. In this way, interview data are inevitably co-created through the interview encounter, yet the interviewee's account is grounded in their existing mental model and own words.

The avoidance of leading is far from being the sole criterion of interview quality. Debates about what constitutes quality in interviews are lively and include considerations of epistemological perspectives (Alvesson, 2003), the characteristics of the interviewer (Brinkmann & Kvale, 2018) and attention to the interactive nature of interviews (Potter & Hepburn, 2005). Roulston (2010) acknowledges that all interviewers need to ask "questions in effective ways" (p.202) as one of four criteria to assess quality and Kvale (1996) summarizes six criteria for assessing interview quality. Nevertheless, the problem of leading is important. The typology contributes a more refined understanding of what it is that makes a question leading by setting out three specific features. This is more informative and specific than the typical advice to avoid leading questions, as detailed in the first part of this paper.

The typology is, however, specific to the wording of questions and statements. It does not address all forms of leading; for example it does not include nonverbal leading (Duncan, Rosenberg, & Finkelstein, 1969) or the influence of interview setting (Potter & Hepburn, 2012). We also acknowledge that the authority given to interviewers (Wang & Yan, 2012) is one of a number of other factors that can affect the likelihood that interviewees will accommodate the interviewer's assumptions. It is relevant to note, nevertheless, that facilitating interviewees to respond from their own frame of reference can mitigate the imbalance of power in the interview.

In order to be practical and accessible we have simplified the theory informing the typology. Linguists have gone into much more detail about the form and nature of presuppositions, for example Beaver & Guerts (2014). The typology is also, of course, dependent on assumptions made by the theories that inform it, which are themselves subject to critique. For example, Lakoff and

Johnson's (1980) perspective on metaphor is debated in the field of cognitive linguistics (Vervaeke & Kennedy, 2004).

One limitation in relation to the typology is that we cannot say for certain it is exhaustive. We have searched assiduously for exceptions by examining extracts from transcripts in authoritative qualitative research publications (Brinkmann & Kvale, 2014; Marshall & Rossman, 2015; Moustakas, 1994) and in studies for which we had access to complete transcripts (Cairns-Lee, 2017; Linder-Pelz & Lawley, 2015; Tosey et al., 2014). Based on this we are confident that these three features are the most prevalent in identifying leading questions.

Turning to the cleanness rating method, we acknowledge that assignment to one of its categories can be a complex task that needs to take into account the contextual, relational and discursive nature of interviews. For example, we have found it relatively straightforward to allocate interviewer interventions to the 'classically clean', 'clean repeat' and 'other' categories. More discernment is required to assign questions at the boundaries between 'contextually clean', 'mildly leading' and 'strongly leading'. An accumulation of individual mildly leading questions may, for example, be sufficient to warrant a strongly leading designation.

The distinctions in the cleanness rating and the adjustments that make questions more or less leading can be subtle. Some familiarity with, and training in, clean language may be advisable. This represents both an advantage and a limitation. The rating is accessible but not mechanistic; yet it needs knowledge, experience and discernment to apply and this requires time and effort. It functions as a stimulus to thought about the kinds of questions posed by a researcher rather than a checklist. Moreover, in our experience, the cleanness rating is most likely to be effective when applied by more than one reviewer (at least one of whom is trained in or well-informed about clean language) as comparison of ratings and discussion with colleagues is likely to improve calibration and help to assign complex cases. Ultimately, we recognize that there is a trade-off between the costs and benefits of using such a method.

The cleanness rating can certainly benefit from further testing and application. The scarcity of interview transcripts in the literature has limited the opportunity to apply the cleanness rating to published studies, especially interviews conducted by interviewers unfamiliar with clean language.

Nevertheless, there is some evidence that the categories of the cleanness rating are robust in that inter-rater reliability is supported. Five projects using clean language as an interview method that employed multiple raters were reviewed by Nehyba and Lawley (2020) for inter-rater reliability. Their analysis demonstrated substantial agreement among raters with an average intraclass correlation coefficient of 0.72 (95% CI). Further research could test the cleanness rating through applying it to various genres of interview.

The cleanness rating has the potential to be used in several ways. Our intention is to offer it primarily as a means of enhancing researcher reflexivity. Johnson and Duberley (2015) highlight the importance of providing a reflexive account of how data are produced in management research “that shows how the voices of the researched are present and meaningful” (p.767). There are educational benefits for the researcher as the rating provides feedback to interviewers that can help them improve their practice. This benefit in particular applies to practitioners as well as scholars. It could be particularly useful to researchers who wish to refine their interview questions during a pilot phase (Young, Rose, Mumby, Benitez-Capistros, Derrick, Finch, Garcia, & al., 2018). Therefore, the method is offered for use by researchers themselves to reflect on their interview questions, not to be externally imposed.

Another potential benefit of the method is that it could be used to report succinctly on confirmability. Limitations on journal article length prevent publication of lengthy transcript extracts. Presenting the results of the rating would be one way to address the influence of the interviewer on the data and the framing of questions advocated by Schaefer & Alvesson (2020) As argued earlier, this is less likely to be relevant to studies in which interviews are regarded as social construction but could be of value to those that aim to understand the subjective perspective of interviewees.

Conclusion

Interviews remain the most pervasive source of data collection in qualitative research. The purpose of this paper has been to enhance understanding of the ways in which questions can influence the authorship of data and the confirmability of findings. The typology of features of leading contributes an improved understanding of the sources of potential bias in interview questions and assists researchers to identify introduced content, presuppositions and evaluation.

Accepted manuscript version of:

Cairns-Lee H, Lawley J, Tosey P. Enhancing Researcher Reflexivity About the Influence of Leading Questions in Interviews. *The Journal of Applied Behavioral Science*. August 2021. doi:10.1177/00218863211037446

The cleanness rating provides a method through which researchers can review the degree to which their questions may have led the interviewee towards the interviewer's own perspective when judged in relation to interviewees' responses. The principal benefit of the rating is to enhance reflexivity. It draws attention to the interviewer's role in the research process in respect of the potential influence of their questions and it can contribute to methodological transparency by enabling researchers to report on confirmability. Our hope, regardless of whether researchers use the typology and cleanness rating method or not, is that this paper increases interviewers' awareness of and ability to manage the potential of interview questions to lead and influence authorship.

References

- Aguinis, H., Ramani, R. S., & Alabduljader, N. (2018). What you see is what you get? Enhancing methodological transparency in management research. *Academy of Management Annals*, 12(1), 83-110. <https://doi.org/10.5465/annals.2016.0011>
- Aguinis, H., & Solarino, A. M. (2019). Transparency and replicability in qualitative research: The case of interviews with elite informants. *Strategic Management Journal*, 40(8), 1291-1315. <https://doi.org/https://doi.org/10.1002/smj.3015>
- Akram, S. (2013). Fully unconscious and prone to habit: The characteristics of agency in the structure and agency dialectic. *Journal for the Theory of Social Behaviour*, 43(1), 45-65. <https://doi.org/10.1111/jtsb.12002>
- Alvesson, M. (2003). Beyond neopositivists, romantics, and localists: A reflexive approach to interviews in organizational research. *The Academy of Management Review*, 28(1), 13-33. <https://doi.org/10.5465/amr.2003.8925191>
- Auger, P., & Woodman, R. W. (2016). Creativity and intrinsic motivation: Exploring a complex relationship. *The Journal of Applied Behavioral Science*, 52(3), 342-366. <https://doi.org/10.1177/0021886316656973>
- Barner, R. (2008). The dark tower: Using visual metaphors to facilitate emotional expression during organizational change. *Journal of Organizational Change Management*, 21(1), 120-137. <https://doi.org/10.1108/09534810810847075>
- Beaver, D. I., & Geurts, B. (2014). Presupposition. In Zalta, E. N. (Ed.), *The Stanford Encyclopedia of Philosophy* (Vol. Winter 2014 Edition). The Metaphysics Research Lab Center for the Study of Language and Information Stanford University. <https://plato.stanford.edu/archives/win2014/entries/presupposition>
- Bluhm, D. J., Harman, W., Lee, T. W., & Mitchell, T. R. (2011). Qualitative research in management: A decade of progress. *Journal of Management Studies*, 48(8), 1866-1891. <https://doi.org/10.1111/j.1467-6486.2010.00972.x>

- Brinkmann, S., & Kvale, S. (2014). *InterViews: Learning the craft of qualitative research interviewing* (3rd ed.). SAGE Publications.
- Brinkmann, S., & Kvale, S. (2018). *Doing interviews* (2nd ed.). SAGE Publications.
- Cairns-Lee, H. M. (2017). *An exploration of leadership and its development through the inner worlds of leaders using metaphor* [Thesis, University of Surrey]. Guildford, Surrey. <http://epubs.surrey.ac.uk/id/eprint/845188>
- Cassell, C., Bishop, V., Symon, G., Johnson, P., & Buehring, A. (2009). Learning to be a qualitative management researcher. *Management Learning*, 40(5), 513-533. <https://doi.org/10.1177/1350507609340811>
- Cassell, C., & Lee, B. (2012). Driving, steering, leading, and defending: Journey and warfare metaphors of change agency in trade union learning initiatives. *Journal of Applied Behavioral Science*, 48(2), 248-271. <https://doi.org/10.1177/0021886312438861>
- Chreim, S., & Tafaghod, M. (2012). Contradiction and sensemaking in acquisition integration. *The Journal of Applied Behavioral Science*, 48(1), 5-32. <https://doi.org/10.1177/0021886311403311>
- Coghlan, D. (2021). Edgar Schein on change: Insights into the creation of a model. *The Journal of Applied Behavioral Science*, 57(1), 11-19. <https://doi.org/10.1177/0021886320924029>
- Conklin, T. A. (2014). Phenomenology redux: Doing phenomenology, becoming phenomenological. *Organization Management Journal*, 11(2), 116-128. <https://doi.org/10.1080/15416518.2014.929935>
- Coruzzi, C. A. (2020). Leading change with intelligence: The power of diagnosis in creating organizational renewal. *The Journal of Applied Behavioral Science*, 56(4), 420-436. <https://doi.org/10.1177/0021886320953982>
- Creswell, J. W., & Poth, C. N. (2018). *Qualitative Inquiry and Research Design: Choosing Among Five Approaches* (4th ed.). SAGE Publications.
- Crowne, D. P., & Marlowe, D. (1964). *The approval motive*. John Wiley.
- Cunliffe, A. L. (2003). Reflexive inquiry in organizational research: Questions and possibilities. *Human Relations*, 56(8), 983-1003. <https://doi.org/10.1177/00187267030568004>
- Doyle, N., & McDowall, A. (2015). Is coaching an effective adjustment for dyslexic adults? *Coaching: An International Journal of Theory, Research and Practice*, 8(2), 154-168. <https://doi.org/10.1080/17521882.2015.1065894>
- Duncan, S., Rosenberg, M., & Finkelstein, J. (1969). The paralinguistic of experimenter bias. *Sociometry*, 32(2), 207-219. <https://doi.org/10.2307/2786264>
- Duží, M., & Číhalová, M. (2015). Questions, answers, and presuppositions. *Computación y Sistemas*, 19(4), 647-659. <https://doi.org/10.13053/CyS-19-4-2327>
- Eberle, T. S. (2014). Phenomenology as a research method. In Flick, U. (Ed.), *The SAGE handbook of qualitative data analysis* (pp. 184-202). SAGE Publications. <https://doi.org/10.4135/9781446282243>
- Englander, M. (2012). The interview data collection in descriptive phenomenological human scientific research. *Journal of Phenomenological Psychology*, 43(1), 13-35. <https://doi.org/10.1163/156916212X632943>

- Foldy, E. G. (2012). Something of collaborative manufacture: The construction of race and gender identities in organizations. *The Journal of Applied Behavioral Science*, 48(4), 495-524. <https://doi.org/10.1177/0021886312440041>
- Foldy, E. G., & Buckley, T. R. (2017). Reimagining cultural competence: Bringing buried dynamics into the light. *The Journal of Applied Behavioral Science*, 53(2), 264-289. <https://doi.org/10.1177/0021886317707830>
- Ford, J. K., Lauricella, T. K., Van Fossen, J. A., & Riley, S. J. (2020). Creating energy for change: The role of changes in perceived leadership support on commitment to an organizational change initiative. *The Journal of Applied Behavioral Science*, February(0), 1-21. <https://doi.org/10.1177/0021886320907423>
- Gibbs, R. W., Jr. (2008). Metaphor and thought. In Gibbs, R. W., Jr. (Ed.), *The Cambridge handbook of metaphor and thought* (pp. 3-14). Cambridge University Press. <https://doi.org/10.1017/CBO9780511816802.002>
- Gill, M. J. (2014). The possibilities of phenomenology for organizational research. *Organizational Research Methods*, 17(2), 118-137. <https://doi.org/10.1177/1094428113518348>
- Gioia, D. (2021). A systematic methodology for doing qualitative research. *The Journal of Applied Behavioral Science*, 57(1), 20-29. <https://doi.org/10.1177/0021886320982715>
- Goddard, C., & Wierzbicka, A. (2014). *Words and meanings: Lexical semantics across domains, languages and cultures*. Oxford University Press. <https://doi.org/10.4000/lexis.3514>
- Goldstein, G., & Kenen, R. (1988). "Internal dialogue" in a "normal" population: The implications for health promotion. *Health Promotion International*, 3(3), 249-257. <https://doi.org/10.1093/heapro/3.3.249>
- Gover, L., & Duxbury, L. (2018). Inside the onion: Understanding what enhances and inhibits organizational resilience. *The Journal of Applied Behavioral Science*, 54(4), 477-501. <https://doi.org/10.1177/0021886318797597>
- Groppe-Wegener, A. (2015). Design tasks beyond the studio. 3rd International Conference for Design Education Researchers 28-30 June 2015, Aalto University.
- Grove, D. J., & Panzer, B. I. (1989). *Resolving traumatic memories: Metaphors and symbols in psychotherapy*. Irvington Publishers.
- Gubrium, J. F., & Holstein, J. A. (2011). *Handbook of interview research: Context and method*. SAGE Publications. <https://doi.org/10.4135/9781412973588>
- Hammersley, M. (2004). Reflexivity. In Lewis-Beck, M., Bryman, A., & Liao, T. F. (Eds.), *The SAGE Encyclopedia of Social Science Research Methods* (pp. 934). SAGE Publications. <https://doi.org/10.4135/9781412950589>
- Hammersley, M. (2017). Interview data: A qualified defence against the radical critique. *Qualitative Research*, 17(2), 173-186. <https://doi.org/10.1177/1468794116671988>
- Hatch, M. J., & Yanow, D. (2008). Methodology by metaphor: Ways of seeing in painting and research. *Organization Studies*, 29(1), 23-44. <https://doi.org/10.1177/0170840607086635>

- Higgs, M., & Rowland, D. (2011). What does it take to implement change successfully? A study of the behaviors of successful change leaders. *Journal of Applied Behavioral Science*, 47(3), 309-335. <https://doi.org/10.1177/0021886311404556>
- Hubbard, A. W. (1950). Phrasing questions. *Journal of Marketing*, 15(1), 48-56. <https://doi.org/10.1177/002224295001500106>
- Hurlburt, R. T., & Heavey, C. L. (2006). *Exploring inner experience: The descriptive experience sampling method*. John Benjamins.
- Husserl, E. (1982). *Ideas pertaining to a pure phenomenology and to a phenomenological philosophy - first book: General introduction to a pure phenomenology* (Kersten, F., Trans.). Kluwer Academic Publishers.
- Isabella, L. A. (1990). Evolving interpretations as change unfolds: How managers construe key organisational events. *Academy of Management Journal*, 33(1), 7-14. <https://doi.org/10.2307/256350>
- Jacobs, C. D., Oliver, D., & Heracleous, L. (2013). Diagnosing organizational identity beliefs by eliciting complex, multimodal metaphors. *Journal of Applied Behavioral Science*, 49(4), 485-507. <https://doi.org/10.1177/0021886313485999>
- Jensen, D. (2008). Confirmability. In Given, L. (Ed.), *The SAGE Encyclopedia of Qualitative Research Methods* (pp. 112). SAGE Publications. <https://doi.org/10.4135/9781412963909.n60>
- Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination and reason*. University of Chicago Press.
- Johnson, P., & Duberley, J. (2003). Reflexivity in management research. *Journal of Management Studies*, 40(5), 1279-1303. <https://doi.org/10.1111/1467-6486.00380>
- Johnson, P., & Duberley, J. (2015). Inductive praxis and management research: Towards a reflexive framework. *British Journal of Management*, 26(4), 760-776. <https://doi.org/10.1111/1467-8551.12103>
- King, N., Horrocks, C., & Brooks, J. (Eds.). (2019). *Interviews in qualitative research* (2nd ed.). SAGE Publications.
- Kovecses, Z. (2002). *Metaphor: A practical introduction*. Oxford University Press.
- Kram, K. E., Wasserman, I. C., & Yip, J. (2012). Metaphors of identity and professional practice: Learning from the scholar-practitioner. *Journal of Applied Behavioral Science*, 48(3), 304-341. <https://doi.org/10.1177/0021886312439097>
- Kvale, S. (1996). *Interviews: An introduction to qualitative research interviewing*. SAGE Publications.
- Kvale, S. (2007). *Doing interviews*. SAGE Publications. <https://doi.org/10.4135/9781849208963>
- Lakoff, G. (1993). The contemporary theory of metaphor. In Ortony, A. (Ed.), *Metaphor and thought*. Cambridge University Press.
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.
- Langley, A., & Meziani, N. (2020). Making interviews meaningful. *Journal of Applied Behavioral Science*, 56(3), 370-391. <https://doi.org/10.1177/0021886320937818>
- Lawley, J. (2017). Clean language interviewing: Making qualitative research interviews verifiable. In Svec, V., Nehyba, J., & Svojanovsk, P. (Eds.), *Becoming a teacher: The*

- dance between tacit and explicit knowledge* (Vol. Education in theory and practice Volume 4, pp. 115-122). Masarykova University.
- Lawley, J., & Tompkins, P. (2000). *Metaphors in mind: Transformation through symbolic modelling*. The Developing Company Press.
- LeVasseur, J. J. (2003). The problem of bracketing in phenomenology. *Qualitative Health Research*, 13(3), 408-420. <https://doi.org/10.1177/1049732302250337>
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE Publications.
- Linder-Pelz, S., & Lawley, J. (2015). Using Clean Language to explore the subjectivity of coachees' experience and outcomes. *International Coaching Psychology Review*, 10(2), 161-174.
- Marshall, C., & Rossman, G. B. (2015). *Designing qualitative research* (6 ed.). SAGE Publications.
- Maxton, P. J., & Bushe, G. R. (2018). Individual cognitive effort and cognitive transition during organization development. *The Journal of Applied Behavioral Science*, 54(4), 424-456. <https://doi.org/10.1177/0021886318795391>
- Mishler, E. G. (2009). *Research interviewing: Context and narrative*. Harvard University Press.
- Moustakas, C. (1994). *Phenomenological research methods*. SAGE Publications.
- Nehyba, J., & Lawley, J. (2020). Clean language interviewing as a second-person method in the science of consciousness. *Journal of Consciousness Studies*, 27(1-2), 94-119.
- Nehyba, J., & Svojanovský, P. (2017). Clean language as a data collection tool. In Svec, V., Nehyba, J., & Svojanovský, P. (Eds.), *Becoming a teacher: The dance between tacit and explicit knowledge* (pp. 130-147). Masarykova University.
- O'Broin, A., & Palmer, S. (2010). Exploring key aspects in the formation of coaching relationships: Initial indicators from the perspective of the coachee and the coach. *Coaching: An International Journal of Theory, Research and Practice*, 3(2), 124-143. <https://doi.org/10.1080/17521882.2010.502902>
- Olson, K., & Bilgen, I. (2011). The role of interviewer experience on acquiescence. *Public Opinion Quarterly*, 75(1), 99-114. <https://doi.org/10.1093/poq/nfq067>
- Oswick, C., Grant, D., & Oswick, R. (2020). Categories, crossroads, control, connectedness, continuity, and change: A metaphorical exploration of COVID-19. *The Journal of Applied Behavioral Science*, 56(3), 284-288. <https://doi.org/10.1177/0021886320936257>
- Owen, I., Rory. (1996). Clean language: A linguistic-experiential phenomenology. In Tymieniecka, A.-T. (Ed.), *Analecta Husserliana: The yearbook of phenomenological research* (Vol. 48, pp. 271-297). Kluwer Academic.
- Patton, M. Q. (2015). *Qualitative research and evaluation methods* (4 ed.). SAGE Publications.
- Payne, S. L. B. (1951). *The art of asking questions*. Princeton University Press.
- Petriglieri, G., Ashford, S. J., & Wrzesniewski, A. M. Y. (2019). Agony and ecstasy in the gig economy: Cultivating holding environments for precarious and personalized work identities. *Administrative Science Quarterly*, 64(1), 124-170. <https://doi.org/10.1177/0001839218759646>
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended

- remedies. *Journal of Applied Psychology*, 88(5), 879-903. <https://doi.org/10.1037/0021-9010.88.5.879>
- 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., & Hepburn, A. (2012). Eight challenges for interview researchers. In Gubrium, J. F., Holstein, J. A., Marvasti, A. B., & McKinney, K. D. (Eds.), *The SAGE handbook of interview research: The complexity of the craft* (2 ed.). SAGE Publications. <https://doi.org/10.4135/9781452218403>
- Pratt, M. G., Kaplan, S., & Whittington, R. (2020). Editorial essay: The tumult over transparency: Decoupling transparency from replication in establishing trustworthy qualitative research. *Administrative Science Quarterly*, 65(1), 1-19. <https://doi.org/10.1177/0001839219887663>
- Qu, Q. S., & Dumay, J. (2011). The qualitative research interview. *Qualitative Research in Accounting and Management*, 8(3), 238-264. <https://doi.org/10.1108/11766091111162070>
- Rapley, T. (2012). The (extra)ordinary practices of qualitative interviewing. In Gubrium, J. F., Holstein, J. A., Marvasti, A. B., & McKinney, K. D. (Eds.), *The SAGE handbook of interview research: The complexity of the craft* (2 ed.). SAGE Publications. <https://doi.org/10.4135/9781452218403>
- Rogers, C. R. (1945). The nondirective method as a technique for social research. *American Journal of Sociology*, 50(4), 279-283. <https://doi.org/10.1086/219619>
- Rogers, K. M., Corley, K. G., & Ashforth, B. E. (2017). Seeing more than orange: Organizational respect and positive identity transformation in a prison context. *Administrative Science Quarterly*, 62, 219-269. <https://doi.org/10.1177/0001839216678842>
- Roulston, K. (2010). Considering quality in qualitative interviewing. *Qualitative Research*, 10(2), 199-228. <https://doi.org/10.1177/1468794109356739>
- Rubin, H. J., & Rubin, L. S. (2005). *Qualitative interviewing: The art of hearing data* (2nd ed.). SAGE Publications.
- Schaefer, S. M., & Alvesson, M. (2020). Epistemic attitudes and source critique in qualitative research. *Journal of Management Inquiry*, 29(1), 33-45. <https://doi.org/10.1177/1056492617739155>
- Schyns, B., Tymon, A., Kiefer, T., & Kerschreiter, R. (2013). New ways to leadership development: A picture paints a thousand words. *Management Learning*, 44(1), 11-24. <https://doi.org/10.1177/1350507612456499>
- Sharma, G., & Good, D. (2013). The work of middle managers: Sensemaking and sensegiving for creating positive social change. *The Journal of Applied Behavioral Science*, 49(1), 95-122. <https://doi.org/10.1177/0021886312471375>
- Sharman, S. J., & Powell, M. B. (2012). A comparison of adult witnesses' suggestibility across various types of leading questions. *Applied Cognitive Psychology*, 26(1), 48-53. <https://doi.org/10.1002/acp.1793>

- Smircich, L., & Morgan, G. (1982). Leadership: The management of meaning. *Journal of Applied Behavioral Science*, 18(3), 257-273. <https://doi.org/10.1177/002188638201800303>
- Sonenshein, S. (2010). We're changing-or are we? Untangling the role of progressive, regressive, and stability narratives during strategic change implementation. *The Academy of Management Journal*, 53(3), 477-512. <https://doi.org/10.5465/amj.2010.51467638>
- Sverdrup, T. E., & Schei, V. (2015). "Cut me some slack": The psychological contracts as a foundation for understanding team charters. *The Journal of Applied Behavioral Science*, 51(4), 451-478. <https://doi.org/10.1177/0021886314566075>
- Tanggaard, L. (2007). The research interview as discourses crossing swords: The researcher and apprentice on crossing roads. *Qualitative Inquiry*, 13(1), 160-176. <https://doi.org/10.1177/1077800406294948>
- Thibodeau, P. H., Hendricks, R. K., & Boroditsky, L. (2017). How linguistic metaphor scaffolds reasoning. *Trends in Cognitive Sciences*, 21(11), 852-863. <https://doi.org/10.1016/j.tics.2017.07.001>
- Tosey, P., Lawley, J., & Meese, R. (2014). Eliciting metaphor through clean language: An innovation in qualitative research. *British Journal of Management*, 25(3), 629-646. <https://doi.org/10.1111/1467-8551.12042>
- Vermersch, P. (2009). Describing the practice of introspection. *Journal of Consciousness Studies*, 16(10-12), 20-57.
- Vervaeke, J., & Kennedy, J. M. (2004). Conceptual metaphor and abstract thought. *Metaphor and Symbol*, 19(3), 213-231. https://doi.org/10.1207/s15327868ms1903_3
- Wang, J. (2006). Questions and the exercise of power. *Discourse & Society*, 17(4), 529-548. <https://doi.org/10.1177/0957926506063127>
- Wang, J., & Yan, Y. (2012). The interview question. In Gubrium, J. F., Holstein, J. A., Marvasti, A. B., & McKinney, K. D. (Eds.), *The SAGE handbook of interview research: The complexity of the craft* (2 ed.). SAGE Publications. <https://doi.org/10.4135/9781452218403>
- Yeo, A., Legard, R., Keegan, J., Ward, K., McNaughton Nicholls, C., & Lewis, J. (2014). In depth interviews. In Ritchie, J., Lewis, J., McNaughton Nicholls, C., & Ormston, R. (Eds.), *Qualitative research practice: A guide for social science students and researchers* (2 ed., pp. 191-192). SAGE Publications.
- Young, J. C., Rose, D. C., Mumby, H., Benitez-Capistros, F., Derrick, C., Finch, T., Garcia, C., & al., e. (2018). A methodological guide to using and reporting on interviews in conservation science research. *Methods in Ecology and Evolution*, 9(1), 10-19. <https://doi.org/10.1111/2041-210X.12828>

Leading Feature	Description	Potential Effect on Authorship
-----------------	-------------	--------------------------------

Introduced Content	Interviewer uses terms that have not previously been used by the interviewee.	Interviewee adopts interviewer's terms.
Presupposition	Can take two forms: a) the <i>structure or syntax</i> of the question presupposes a situation that the interviewee has not previously stated exists; b) the question presupposes a <i>logical relationship</i> between two or more items not specified by the interviewee (often a cause-effect relationship is assumed).	Unless the interviewer's presupposition is rejected, the interviewee may adopt the assumption of the presupposed situation or relationship.
Evaluation	The interviewer expresses an <i>opinion or evaluation, or raises a doubt or objection</i> about something said by the interviewee.	May serve to undermine the interviewee's opinions or confidence in the value of their own experience and thus make them more susceptible to later leading questions and statements.

Table 1: Leading features of interview questions

Cleanness Rating Category	Description of Category
Classically Clean	Categories refer to introduced content, presupposition & evaluation that are defined as Leading Features of Interview Questions in Table 1
Clean Repeat	A question that only makes use of universal constructs together with interviewee content. Such a question will use a variant of classically clean questions (listed in Table 3).
Contextually Clean	<p><i>Topically clean:</i> A question which introduces the interview topic while minimizing superfluous content and presupposition.</p> <p><i>Logically clean:</i> A question (other than classically clean) that remains within the logic of the interviewee's descriptions and does not introduce interviewer content, presupposition or evaluation.</p>
Mildly Leading	A question or statement that suggests or implies an answer or way of answering (due to the presence of introduced content, presupposition or evaluation), but the interviewee's response gives the rater no reason to doubt the authorship of the answers.
Strongly Leading	A question or statement that suggests or implies an answer or way of answering (due to the presence of introduced content, presupposition or evaluation) and the response gives the rater reason to doubt the authorship of the interviewee answers.
Other	A non-leading comment (such as aha, mm-hmm, okay) that encourages the interviewee to continue and indicates the interviewer is paying attention; or a statement/question about the interview process; or a response to an interviewee question during the interview.

Table 2: Definition of the six categories in the cleanness rating

ATTRIBUTES	And what kind of X is that X? And is there anything else about X?
LOCATION	And where/whereabouts is X?
REFLEXIVITY	And how do you know?
METAPHOR	And that's X like what?
RELATIONSHIP	And when X, what happens to Y? And is X the same or different as Y?
SEQUENCE	And then what happens/what happens next? And what happens just before X?
SOURCE	And where does/could X come from?
Where 'X' and 'Y' = interviewee words.	

Table 3: Classically clean questions adapted from Tosey et al., 2014 p. 631