EUROPEAN JOURNAL OF WORK AND ORGANIZATIONAL PSYCHOLOGY
https://doi.org/10.1080/1359432X.2019.1704265

Encouraging the production of narrative responses to past-behaviour interview questions: effects of probing and information

Julie Brosy, Adrian Bangerter and Sandrine Ribeiro

Institute of Work and Organizational Psychology, University of Neuchâtel, Neuchâtel, Switzerland

ABSTRACT

In behavioural interviews, past-behaviour questions invite applicants to tell a story about a past job-related situation. Nevertheless, applicants often do not produce stories on demand, resorting to less appropriate responses. In a sample of real selection interviews (Study 1), only 50% of applicants' responses to past-behaviour questions were indeed stories. We explored two factors that may increase applicants' storytelling tendencies: probing and information about past-behaviour questions. In two experiments simulating selection interviews, we manipulated recruiter probing during the interview (Study 2) and the level of participants' information about the expected answer format of past-behaviour questions (Studies 2 and 3). Probing induced participants to tell more stories and to include more narrative diversity in their stories, but there was no effect of giving participants information or not. More information did help participants to tell less pseudostories (generic descriptions of situations). Analyses of participants' thoughts and emotions experienced during question-answering suggest that finding an appropriate example to narrate is a major problem. Storytelling rate also varied by competency. Findings are relevant for theories of behaviour elicitation in selection situations.

ARTICLE HISTORY
Received 8 October 2018
Accepted 27 November 2019

KEYWORDS
Selection interview; past-behaviour question; storytelling; probing; behaviour elicitation

Introduction

Research in personnel selection has focused heavily on the validity of selection methods (Schmidt & Hunter, 1998), especially for interviews, which have been historically dogged by concerns about predictive validity (Judge, Cable, & Higgins, 2000). Recently, however, a consensus has emerged that selection interviews can attain high levels of validity if they are structured (e.g., Huffcutt & Arthur, 1994; Levashina, Hartwell, Morgeson, & Campion, 2014). Structure involves practices (Campion, Palmer, & Campion, 1997; Huffcutt & Arthur, 1994) designed to influence the content of the interview or evaluation of applicants, and often limits interactivity. Despite the better validity of structured interviews, practitioners remain reluctant to use them (Lievens & De Paepe, 2004). However, one specific type of structured interview, the behavioural interview, may be becoming more popular (Ralston, Kirkwood, & Burant, 2003; Roulin & Bangerter, 2012) and evidences high predictive validity (Huffcutt, Conway, Roth, & Klehe, 2004; Huffcutt, Weekley, Wiesner, Degroot, & Jones, 2001).

In behavioural interviews, job analysis is used to identify relevant competencies. Applicants are then asked so-called past-behaviour questions about their actions in past situations (e.g., *Can you tell me about a time you had to deal with an angry client?*) and their responses are rated to gauge their mastery of those competencies. Past-behaviour questions invite applicants to produce *narrative responses*, i.e., to *tell a story* about the situation, thereby making storytelling skills increasingly important (Ralston et al., 2003). Many studies have investigated applicants' responses in terms of broad categories of impression management tactics in structured interviews (e.g., self-promotion, ingratiation) and their impact on interviewers' evaluations (Ellis, West, Ryan, & DeShon, 2002; Van Iddekinge, McFarland, & Raymark, 2007). However, these studies typically do not analyse the actual discourse produced. As a result, little is known about the prevalence of personal stories in interviews, about applicants' storytelling abilities, about how to help applicants produce better responses using storytelling, or about how to evaluate them (Bangerter, Corvalan, & Cavin, 2014; Levashina et al., 2014; Ralston et al., 2003). For even perhaps the most basic question, how prevalent stories are in response to past-behaviour questions, studies show inconsistent results. Stevens and Kristof (1995) found that many applicants in a field study spontaneously told stories (not necessarily in response to past-behaviour questions). However, Bangerter et al. (2014) found that only 23% of applicants told stories in response to past-behaviour questions.

Thus, there is much scope for helping applicants improve their storytelling performance. These could potentially lead to better functioning of the behavioural interview. If applicants can be trained to produce better stories, they may be able to better communicate the true level of their competencies, thus potentially increasing the predictive and construct validity of the interview. This in turn illustrates a more general theoretical issue of how to elicit trait-expressive behaviours in selection situations (De Kock, Lievens, & Born, 2018). Recent approaches suggest that recruiters' active participation in interactions with applicants may increase the availability and relevance of behavioural cues of job-relevant traits. Thus, recruiters may play a potentially important role in facilitating storytelling.

A major contribution of this paper is to document the suboptimal functioning of past-behaviour questions and to test

CONTACT Adrian Bangerter adrian.bangerter@unine.ch

possible improvements. We first measured how often stories get produced in response to past-behaviour questions in real selection interviews (Study 1). We then conducted two experiments (Studies 2 and 3) exploring factors that may increase storytelling in simulated interviews, where participants played the role of applicants (we hereafter use this label). We manipulated how much *information* applicants received about the responses expected of them as well as whether or not experimenters (who played the role of recruiters; we hereafter use this label) provided cues to applicants about expected responses, i.e., were allowed to engage in *probing*. Our findings suggest explanations for difficulties applicants may experience in answering past-behaviour questions and suggest ways recruiters may help them produce better stories, supporting theories of active cue elicitation in selection situations (De Kock et al., 2018). In the next sections, we define and motivate the role of storytelling and stories in selection interviews, before describing the rationale of our studies.

Storytelling and stories in selection interviews

Stories can be distinguished from storytelling (Mandelbaum, 2013). Storytelling is a collaborative activity involving narrator(s) and an audience. A story is an abstraction from that activity, i.e., a set of propositions describing an episode the narrator(s) have experienced. In interviews featuring past-behaviour questions, recruiters and applicants are participants in a storytelling activity (applicants are narrators and recruiters are the audience). Afterwards, recruiters reconstruct the story as a basis for evaluating applicants. Recommended practices for structured interviewing aim at minimizing recruiters' involvement in the social interaction (Campion et al., 1997; Dipboye, 1994). But such practices may also inhibit productive storytelling activities. Designing recommendations for optimal recruiter behaviour should thus build on properties of both the storytelling activity and its product, the story.

Storytelling as a collaborative activity

Storytelling has been widely studied in communication research (Mandelbaum, 2013; Norrick, 2007). Like any conversation, it is a collaborative activity where participants actively coordinate their actions to progress through it. Launching an episode of storytelling involves making it recognizable as such so that participants understand they are expected to suspend the usual turn-taking rules (Sacks, Schegloff, & Jefferson, 1974), to allow narrators extended rights to the floor. Narrators attempt to secure the interest of audiences to gain licence to launch the telling. Progress also depends on listeners' participation. Listeners produce two types of responses, *generic responses* (e.g., nodding, *mhm*) and *specific responses* (e.g., wincing, exclaiming or otherwise commenting on a specific aspect of the narrative). Without listener responses, stories are told less well and end abruptly (Bavelas, Coates, & Johnson, 2000). Listeners are thus co-narrators of a story. Participants also collaborate (Mandelbaum, 2013) to define a story's meaning (e.g., illustrating the character of a person one is gossiping about) and in ending it (e.g., by recognizing its climax) and moving on to other topics.

In selection interviews, the collaborative nature of storytelling means that it involves both applicants and recruiters. Recruiters who behave interactively and participate in the storytelling activity may ultimately help applicants tell better stories (Ralston et al., 2003). Recruiters who limit their involvement in the interaction to minimize bias (Campion et al., 1997; Huffcutt & Arthur, 1994) may inadvertently inhibit productive storytelling.

Stories and their impact

In selection interviews, stories have many potential effects on recruiters. First, they may illustrate applicants' claims (e.g., by showing how applicants overcame obstacles; Stevens & Kristof, 1995). Their concreteness helps imagine what is recounted (Schank & Berman, 2002). They constitute self-promotion tactics, which create an attractive image and increase hiring recommendations. Furthermore, well-told stories may enhance applicants' distinctiveness and help recruiters remember them better (Stevens & Kristof, 1995). Finally, detailed stories are perceived as difficult to fake, thus increasing applicants' credibility (Bangerter, Roulin, & König, 2012). But how do stories actually unfold in interviews?

Stories are "discourse that is organized more or less chronologically and tells what characters did in a particular situation and what came of their actions. Stories may be very brief or extended, and they may depict scenes, characters, actions, and consequences that vary in richness and detail" (Ralston et al., 2003, p. 11). A story meets three criteria (Bangerter et al., 2014; Norrick, 2007; Ralston et al., 2003). First, it is a description of events about a unique past episode. Second, events are chronologically or causally coherent and linked by temporal markers. Third, it provides information about the where, who, what and how of the episode (the situation, the protagonists, what they did and how they resolved the situation). Several models describe the typical constituent parts of a complete story, e.g., Labov's (1972) narrative structure – abstract, orientation, complicating action, evaluation, resolution, and coda. In selection interviews, the STAR mnemonic – situation, task, action and results (Kessler, 2006) is used by recruiters to evaluate the completeness of applicants' stories.

Difficulties with storytelling in the selection interview

Cognitive difficulties may prevent applicants from producing effective stories during the interview (Bangerter et al., 2014; Ralston et al., 2003). Storytelling involves processes of autobiographical recall and response planning and production (Bradburn, Rips, & Shevell, 1987; Brosy, Bangerter, & Mayor, 2016; Huffcutt, Culbertson, Goebl, & Toidze, 2016). Behavioural questions may create strong memory demands because applicants have to recall a suitable episode. The stressful context of the interview may exacerbate this by increasing cognitive load, impeding effective memory search for a relevant past episode. There may also be differences in how easily competencies can be translated into stories (Bangerter et al., 2014). It may be more difficult to find examples for more mundane, everyday competencies (like *organizing*) than for infrequently used competencies (like *negotiation*).

Difficulties in producing stories in response to past-behaviour questions may also relate to social processes.

Finding an example is compounded by the time pressure inherent to the question-answering process, which is subject to self-presentation goals (Smith & Clark, 1993), especially in job interviews (Brosy et al., 2016). Because excessive delays in initiating a response may lead to negative impressions, applicants experiencing difficulties in choosing an example may choose to start talking faster but perhaps less appropriately (Brosy et al., 2016). They may be more likely to talk about recent or highly available events (Huffcutt et al., 2016), or produce pseudostories (generic descriptions of event types rather than a unique event; Bangerter et al., 2014) which may not serve them best.

A further difficulty in producing stories in response to past-behaviour questions results from the collaborative nature of everyday storytelling activities. Because would-be storytellers try to gain the agreement of their audience before initiating a story (Mandelbaum, 2013), applicants may hesitate to take the initiative (Bangerter & Gosteli-Corvalan, 2017) required to tell a story because they are unsure of its appropriateness (Bangerter et al., 2014; Brosy et al., 2016). Moreover, because a story's progress and denouement depend on listeners' actions (Bavelas et al., 2000), in the behavioural interview, interactive behaviour on the part of recruiters (e.g., probing) may be necessary to encourage productive storytelling.

The current studies

We explored storytelling in applicants' responses to past-behaviour questions, investigating how often applicants produced stories in real interviews, what difficulties they face in responding to past-behaviour questions and testing interventions to facilitate storytelling. Beyond the specific context of behavioural interviewing, these issues relate to the general problem of elicitation of trait-expressive behaviour in personnel selection, or how to create felicitous conditions for observing representative samples of behaviour (stories) to measure traits of interest (competencies). Recent models of "good raters" (De Kock et al., 2018) suggest that recruiters are not just passive observers of applicant behaviour in testing situations, but can actively contribute to expression of relevant cues by interacting with applicants.

From the perspective of behaviour elicitation, it is important to establish the efficacy of past-behaviour questions in eliciting stories. Previous studies suggested applicants have difficulty producing stories (Bangerter et al., 2014), but were based on interviews for an entry-level position created for a validation study. Thus, it is important to ascertain how frequent storytelling is in real interviews. In Study 1, we coded how often past-behaviour questions were asked in a diverse sample of real selection interviews and how often applicants responded by telling stories. We formulated Research Question 1 (RQ1): *How often do applicants produce stories in response to recruiters' past-behaviour questions?*

In Studies 2 and 3, we tested interventions to enhance the production of stories in response to past-behaviour questions. For ethical and internal validity reasons (Lonati, Quiroga, Zehnder, & Antonakis, 2018; Podsakoff & Podsakoff, in press), we used experimental methods (simulated interviews). We manipulated two independent variables.

The first variable (Studies 2 and 3) is applicants' level of *information* before the selection interview. According to trait activation theory (Lievens, Schollaert, & Keen, 2015), this constitutes a *general* situational characteristic that moderates the effect of possessing a trait on the activation of a trait-expressive behaviour. We gave applicants information just before the interview about the type of questions they would be asked (i.e., past-behaviour questions), the competencies they would be asked about or the expected response type, allowing them to use this information to prepare their responses. This manipulation allows investigating the role of cognitive factors like memory in recalling appropriate examples. More precisely, we expected the manipulation of information would allow applicants to search in memory for a relevant past episode in a context less stressful than during the interview and with less time pressure inherent to the question-answering process (Huffcutt et al., 2016) and to facilitate response planning and production. This leads to Hypothesis 1 (H1): *Information increases production of stories*. We also expected information to affect the production of pseudostories (Bangerter et al., 2014). Pseudostories are a common response type where applicants describe generic sets of events (typically in the present tense, e.g., *I usually organize my time by setting priorities*) instead of a unique event. Because information alerts applicants to the fact that a story is an expected response (and because applicants have time to find an example to narrate, the information should also reduce the production of pseudostories). We thus formulate Hypothesis 1b (H1b): *Information decreases production of pseudostories*. The information allows targeted preparation, increasing the potential to recall relevant episodes. We thus expected it to increase the narrative diversity of stories, that is, the range of elements like the situation, task, actions or results that applicants would talk about. This leads to Hypothesis 2 (H2): *Information increases narrative diversity of stories*.

The second variable (manipulated in Study 2) is *probing*. We manipulated recruiters' interactive behaviour during the interview by either allowing them to produce probes or not. According to trait activation theory (Lievens et al., 2015), this manipulation constitutes a specific intervention on the part of the recruiter in the interview situation. It allows investigating the role of social factors like the collaborative nature of storytelling. This manipulation responds to calls for more research on the effects of probing (De Kock et al., 2018; Levashina et al., 2014). We expected that recruiter probing would lead applicants to tell stories more often as well as tell more detailed and complete stories, by alerting them to missing information in their response (like insufficient descriptions of actions or results). This leads to the formulation of Hypothesis 3 (H3): *Probing increases applicants' production of stories*, and Hypothesis 4 (H4): *Probing increases the narrative diversity of stories.*

Beyond the production and narrative diversity of stories, we were interested in what psychological processes mediate responding to past-behaviour questions. As described above, it is currently unclear what role is played by cognitive difficulties in recalling an appropriate episode to narrate (Huffcutt et al., 2016) or hesitations about the appropriateness of extended storytelling. Interventions to elicit storytelling

responses would benefit from knowledge about these processes. We thus formulate Research Question 2 (RQ2): *What are the cognitive and emotional processes associated with answering past-behaviour questions?*

Finally, following previous research that found differences in storytelling depending on the competency measured (Bangerter et al., 2014), it seems important to assess the effects of competency on storytelling. This may have practical implications for recruiters. We thus formulate Research Question 3 (RQ3): *Does the production of stories depend on the type of competency measured?*

Study 1

Study 1 investigated RQ1 (*How often do applicants produce stories in response to recruiters' past-behaviour questions?*) using a sample of real selection interviews, thus increasing ecological validity relative to previous studies (Bangerter et al., 2014). It is important to establish that storytelling responses are indeed problematic in order to ground the relevance of any experimental manipulations to increase storytelling.

Methods

Participants

Eighty selection interviews were audiorecorded in two organizations in French-speaking Switzerland. Each interview involved an applicant and two or three recruiters. Applicants (62.5% women, *M*age = 33.95 years) had experienced on average 5.7 interviews (*SD* = 5.2) and had on average 13.07 years (*SD* = 11.5) of professional experience, 25.3% of them had managerial experience and 12.6% had a university diploma. Thirty-one recruiters participated (35.5% of women, Mage = 45.10 years). They conducted on average 22.23 interviews per year. Three were from the HR department and 28 were line managers.

Procedure

We contacted the human resources (HR) department of both organizations to obtain their agreement to participate. They then contacted the other participants. Non-HR recruiters (e.g., line managers) were informed about the study when they selected applicants to interview (76% agreed to participate). They received an information letter and a consent form. Applicants were informed about the study when they were called for an interview (89% agreed to participate). They received an information letter, a consent form and a questionnaire about sociodemographic data and personality. The first author attended and audiorecorded all interviews. Afterwards, applicants filled out a second questionnaire about their use of impression management. Recruiters filled out a questionnaire about their perception of applicant impression management, a hiring recommendation and sociodemographic data. Participants' questions were answered, and they were thanked for their participation. Interviews were transcribed word-for-word as a basis for coding.

Coding

Past-behaviour questions. We coded past-behaviour questions asked by recruiters. A past-behaviour question was calibrated to a competency and asked applicants to talk about a past situation they experienced and ideally where they performed well on the desirable competence (Bangerter et al., 2014).

Stories. We coded the presence or absence of stories for the responses to each past-behaviour question (double-coding of 20 interviews, Cohen's kappa = .76). A story was defined as "a set of events related to a unique past episode, characterized by a unity of time or action, with constituents often linked by temporal markers (e.g., *then*)" (Bangerter et al., 2014, p. 6).

Results and discussion

Past-behaviour questions occurred in 27 interviews (42 questions total). Applicants produced 21 stories in response, meaning that they responded appropriately only 50% of the time. While the rate of storytelling is higher than in previous studies, past-behaviour questions clearly are not optimal techniques for eliciting stories. Indeed, that a behaviour elicitation technique fails to elicit relevant cues half of the time is grounds for serious concern. Thus, results confirm the need to improve understanding of factors enhancing applicant storytelling.

Study 2

Study 2 investigated H1 (effect of information on story production), H2 (effect of probing on narrative diversity of stories), H3 (effect of probing on story production), H4 (effect of probing on narrative diversity of stories) and RQ3 (storytelling production by competency). Hypotheses were tested in a simulated selection interview, where the experimenter (the third author) played the role of the recruiter and participants played the role of applicants. The experiment featured a 2 × 2 between-subjects design *(information* X *probing)*. Because prior research documented links between individual differences and storytelling, we measured participants' gender, age, personality, ability to identify criteria (ATIC) (Kleinmann et al., 2011), general mental ability (GMA) and work and interview experience.

Methods

Participants

We recruited 60 French-speaking students via personal contacts of the experimenter (25 women, *Mage* = 24.20, *SDage* = 3.29). A majority (55) already had a first job and had experienced at least one selection interview (56). They each received 10 Swiss francs for participating and an additional prize of 300 Swiss francs was awarded for the best performance during the interview. The prize was designed to motivate applicants to perform well, i.e., to avoid them simply "going through the moves" in participating.

Independent variables

Information. Participants received written instructions 5 minutes before the interview to prepare themselves. In the

uninformed condition, they received the following instructions: *You have 5 minutes to prepare yourself for the selection interview. Try to imagine questions the recruiter could ask you*, with no information about the type of questions they would be asked. In the *informed* condition, they received the following instructions: *You have 5 minutes to prepare yourself for the selection interview. During this interview, the recruiter will ask you past-behaviour questions, i.e., questions you need to respond to by describing a behaviour. Tell what you did in this specific situation. Example of past-behaviour question: Could you describe a situation where you felt stressed and how you managed it?*

Probing. In the *no probing* condition, the experimenter asked the four past-behaviour questions without any probes contingent on participants' responses. In the *probing* condition, she chose between pre-established probes, depending on the participant's response and following an inappropriate or incomplete response. Following a response not mentioning a specific situation, she asked *Can you give me an example of a situation?* Following an incomplete story, e.g., mention of a specific situation but no action, she asked *Can you tell me a little bit more about what you did in this situation?* or *Can you tell me more about it?* The experimenter used a maximum of two probes per question.

Procedure

Participants were informed the study was about interactions in selection interviews and that they would play the role of applicants. A few days before, they received three job ads (*business analyst, company lawyer* or *HR specialist*) and chose one. All three in fact required the same competencies: *adaptation, autonomy, team work* and *organizing*. Upon arriving for the experiment, they read and signed a consent form. Then, they received the instructions, including the reminder of the prize, and had 5 minutes to prepare themselves. The interview began with two introductory questions to help participants immerse themselves in the simulation. Then, they were asked four past-behaviour questions (with or without probing, depending on the condition). An example question (for the adaptation competency) was *Can you describe how you managed a situation where a sudden and urgent event occurred while you were doing a routine task?* The interview was videotaped. Afterwards, participants filled out a manipulation check and provided individual difference data before debriefing and payment. After the study, the experimenter determined the best-performing participant to award the prize to. Interviews were transcribed word-for-word in preparation for coding.

Measures

Manipulation check. We created three items (e.g., *To what extent did the experimenter take part in the interaction during the interview?*) to check if the probing manipulation was perceived by participants. Participants responded via 5-point Likert scales from 1 (e.g., *not at all*) to 5 (e.g., *very much*), Cronbach's alpha = .83. We computed a total score by participant from 3 to 15. Participants in the probing condition (M = 12.67, SD = 2.29) perceived the experimenter as significantly more interested and participative than in the no-probing condition (M = 9.03, SD = 2.63), $F(1,59) = 32.5$, $p < .001$.

Gender. We measured participant gender (0 = men, 1 = women).

Interview experience. We asked participants how many selection interviews they had experienced.

Professional experience. We asked participants how many years of professional experience they had.

General mental ability (GMA). We assessed GMA with the Wonderlic Personnel Test (Wonderlic, 2001). Participants had 12 minutes answer a maximum of 50 questions. The score is the number of correct responses.

ATIC. We measured the ability to identify criteria (ATIC) with eight items. Instructions were to identify the competencies measured by past-behaviour questions. An example of the autonomy competency is *Can you describe a situation where you took an initiative and achieved it?* Participants' responses were coded as either incorrect or correct by two judges, allowing for computation of interrater agreement. Cohen's kappa was high, varying from .78 to 1 depending on the item. We computed a total score of correct identifications (1–8) by participant (Kleinmann et al., 2011).

Personality. We measured personality with 30 items measuring the five traits of the NEO PI-R (Costa & McCrae, 1992): neuroticism (alpha = .76), extraversion (alpha = .76), openness (alpha = .65), agreeableness (alpha = .59 with one item removed) and conscientiousness (alpha = .86). Items were opposite pairs, an example for extraversion is *social – solitary*. Items were answered on a scale where 1 = *very much*, 2 = *pretty much*, 3 = *quite*, 4 = *quite*, 5 = *pretty much* and 6 = *very much*.

Stories. We coded the presence or absence of stories for the responses of each past-behaviour question, before and after probing (double-coding of 32 responses, Cohen's kappa = .87) and computed the number of responses featuring stories (between 0 and 4) by participant.

Increase of stories. We computed the presence or absence of an increase of stories for participants in the *probing* condition. An increase was counted when there was no story before the probe and a story appeared after it. We excluded one participant in the *probing* condition who was not asked any probes (experimenter error).

Narrative elements. We coded the presence or absence of three narrative elements contained in stories: Situation, Task/Action, and Results (Kessler, 2006), both before and after probing (double-coding of 32 responses, Cohen's kappa between .63 and .76).

Increase of narrative elements. We computed the presence or absence of an increase of each narrative element for participants in the *probing* condition. An increase was counted when the narrative element was not mentioned before a probe and appeared after it.

Results

Descriptive results appear in Table 1. We found no significant correlations between the initial story production rates, the increase of stories and the production of narrative elements and gender, age, personality, ATIC, GMA and work and interview experience. Therefore, we did not use these variables in further analyses.

Hypotheses 1 and 3: effects of information and probing on storytelling

We ran a 2 × 2 between-subjects ANOVA (*information* X *probing*) predicting the initial level of production of stories. Before any probes, participants produced stories 56% of the time. We found no significant effect of *information,* $F(1,56) = 1.46$, $p = .23$, $d = 0.44$, on participants' production of stories. Therefore, H1 was not supported. There was no significant effect of *probing,* $F(1,56) = 2.71$, $p = .106$, $d = 0.60$, and no significant interaction effect, $F(1,56) = .59$, $p = .446$, $d = 0.28$ (neither effects were expected to be significant on the initial production of stories). To test H3 we then focused on the *probing* condition. After probing, participants produced stories 85% of the time. The overall increase was significantly different from 0, $t(28) = 5.51$, $p < .001$, 95% CI [.61, 1.32], $d = 1.03$. Probing induced participants to tell significantly more stories for three competencies out of four (see Table 2, only autonomy did not increase; the initial production rate of stories for autonomy may have been too high for probing to increase it). Overall, therefore, H3 was supported.

Research question 3: storytelling by competency

We conducted pairwise *t*-tests between competencies (Bonferroni correction; $p = .05/6 = .008$) to ascertain whether stories are more easily produced for some competencies than others. Autonomy ($M = .72$, $SD = .45$) was significantly different from teamwork ($M = .47$, $SD = .50$), $t(59) = 3.23$, $p = .002$, $d = 0.83$, and from organizing ($M = .40$, $SD = .49$), $t(59) = 3.77$, $p < .001$, $d = 0.97$. Other comparisons were not significant. The adaptation was not significantly different from autonomy $t(59) = 1.09$, $p = .279$, $d = 0.28$, organizing, $t(59)$, 2.69, $p = .009$, $d = 0.69$, or teamwork $t(59) = 2.32$, $p = .024$, $d = 0.60$. Teamwork was not significantly different from organizing, $t(59) = .78$, $p = .437$, $d = 0.20$. Thus, participants produced significantly more stories when asked about adaptation and autonomy than teamwork and organization (Table 2).

Hypotheses 2 and 4: effects of information and probing on narrative elements in stories

Before any probes, situations were the most frequent narrative element mentioned (35% of the time), following by tasks/actions (31%) and results (15%) (see Table 3). We ran a 2 × 2 between-subjects ANOVA (*information* by *probing*) predicting the narrative elements initially produced. Participants in the *informed* condition did not differ from participants in the *uninformed* condition, $F(1, 60) = .89$, $p = .349$, $d = 0.34$. Before any probes, participants in the *probing* condition differed from participants in the *no probing* condition, $F(1, 60) = 20.49$, $p < .001$, $d = 1.65$. The interaction effect was not significant, $F(1, 60) = .002$, $p = .961$, $d = 0.02$. *Information* did not influence narrative diversity in participants' stories. Therefore, H2 was not supported. We then focused on the *probing* condition. The overall difference was significantly different from 0, $t(28) = 8.85$, $p < .001$, 95% CI [3.05, 4.88], $d = 1.62$. Probing induced participants to include more narrative diversity in their stories by increasing talk about actions and results (see Table 3). Therefore, H4 was supported.

Note that participants in the *probing* condition included more overall narrative elements in their stories before any probes than participants in the *no probing* condition. This unexpected difference could be a learning effect: after one or two questions featuring probes, participants could have been alerted to the narrative elements expected and adapted their subsequent initial responses accordingly. We tested for this in the probing condition by looking at the evolution of narrative elements present in participants' responses to each question before probing. With Bonferroni correction ($p = .05/6 = .008$), results showed that participants actually mentioned less narrative elements in response to Question 3 (teamwork, $M = .87$, $SD = 1.14$), and Question 4 (organizing, $M = .77$, $SD = .94$) than in response to Question 2 (autonomy, $M = 1.83$, $SD = .95$), respectively, $t(29) = 3.92$, $p < .001$, $d = 1.43$, and $t(29) = 5.41$, $p < .001$, $d = 1.98$. There were no significant differences in the narrative elements mentioned between responses to Question 1 (adaptation $M = 1.30$, $SD = 1.12$) and to Question 2, $t(29) = 2.19$, $p = .036$, $d = 0.80$, Question 3, $t(29) = 1.78$, $p = .085$, $d = 0.65$, and Question 4, $t(29) = 1.98$, $p = .058$, $d = 0.72$, and between responses to Questions 3 and 4, $t(29) = .42$, $p = .676$, $d = 0.15$. Thus, participants in the *probing* condition did not increase the narrative elements in their responses as they progressed through the interview, speaking against a potential learning effect. It is unclear what caused this difference between conditions.

Discussion

The main aim of this study was to test the effects of two variables, *information* and *probing*, on participants' production of stories (H1 and H3) and the narrative elements contained in stories (H2 and H4). *Information* did not affect the production of stories (rejecting H1), but *probing* had a significant positive effect (supporting H3). *Information* did not affect the narrative elements contained in stories (rejecting H2), but *probing* had a significant positive effect on the narrative elements contained in the stories (supporting H4).

Results support the idea that recruiters' probing behaviour influences the way applicants respond to past-behaviour questions during selection interviews. Recruiters may have a key role to play in improving the quality of applicants' responses, because their probes increase applicants' storytelling tendencies as well as the narrative elements those contain.

We found differences in storytelling depending on competency (RQ3). When participants were asked about adaptation and autonomy, they produced more stories than when asked about teamwork and organizing. This may reflect an order effect: adaptation and autonomy were the two first competencies asked about. Participants may have had trouble finding more than two examples to recount and may have been reluctant to use the same examples more than once. On the other hand, the differences may be due to the fact that the questions

Table 1. Means (*M*), standard deviations (*SD*) and correlations for main study variables (Study 2).

	Variable	M	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	Gender	0.42	0.50	-																	
2	Age	24.20	3.29	.04	-																
3	Interview experience	0.93	0.25	.23	.10	-															
4	Professional experience	0.92	0.28	.26*	.19	.40**	-														
5	Openness	4.27	0.60	.00	.17	−.01	.22	-													
6	Conscientiousness	4.51	0.98	.31*	.03	−.03	.34**	.46**	-												
7	Extraversion	4.17	0.78	.29*	−.09	−.06	−.01	.29*	.19	-											
8	Agreeableness	24.25	3.05	−.04	−.07	−.24	−.08	.29*	.33*	−.04	-										
9	Neuroticism	2.71	1.07	.10	.19	−.06	−.03	−.03	−.06	.11	.12	-									
10	GMA	22.80	5.65	−.04	.21	−.03	−.09	−.11	−.41**	.08	.00	.20	-								
11	ATIC	2.37	1.33	−.03	.05	−.08	−.33*	−.02	−.14	.00	.05	.01	.30*	-							
12	Information condition	0.50	0.50	.03	.20	.13	.06	−.14	−.21	−.09	−.31*	.20	.21	.23	-						
13	Probing condition	0.50	0.50	−.10	.05	.13	−.06	.03	.00	.12	.11	.12	.19	−.05	.00	-					
14	Stories (before probing)	2.22	1.19	−.04	−.24	.22	.06	.11	−.16	−.01	−.05	.08	.14	.22	.16	.21	-				
15	Stories (after probing)	3.38	0.94	−.40*	−.50**	−.13	−.11	−.11	−.13	−.20	.32	.15	−.09	.18	−.25	n/a	.57**	-			
16	Increase in stories	0.97	0.94	−.20	−.09	−.21	−.14	.09	.16	−.13	.27	.27	−.07	.09	−.26	n/a	−.62**	.26	-		
17	Narrative elements (before)	3.25	2.97	.01	−.07	.20	−.04	.02	−.23	.09	−.12	.07	.26*	.21	.11	.52**	.63**	.40*	−.53**	-	
18	Narrative elements (after)	3.33	3.88	−.15	−.13	.06	−.15	−.08	−.07	.03	.17	.16	.20	.02	−.02	n/a	.32*	.71**	.18	.60**	-
19	Difference narrative elements	1.98	2.64	−.15	−.11	−.03	−.03	−.08	.04	.00	.21	.22	.15	−.04	−.07	n/a	.14	.54**	.56**	.26*	.84**

Gender: 0 = Men, 1 = Women. GMA: General mental ability. ATIC: Ability to identify criteria. Information condition: 0 = uninformed, 1 = informed. Probing condition: 0 = no probing, 1 = probing. n/a: no value because values are only defined in the probing condition.

* $p < .05$; ** $p < .01$; *** $p < .001$.

Table 2. Mean proportions of production of stories and increase by competency (Study 2).

Variable	Before probing (All, N = 60)	Before probing (Probing condition, N = 29)	After probing (Probing condition, N = 29)	Increase (Probing condition, N = 29)
Adaptation	.63	.66	.86	.21*
Autonomy	.72	.86	.97	.10
Teamwork	.47	.48	.90	.41***
Organizing	.40	.48	.66	.24**
Overall	.56	.62	.85	.23***

* $p < .05$; ** $p < .01$; *** $p < .001$.

Table 3. Proportion of narrative elements and difference by element (Study 2).

Variable	Before probing (All, N = 60)	Before probing (Probing condition, N = 29)	After probing (Probing condition, N = 29)	Difference (Probing condition, N = 29)
Situation	.35	.55	.36	−.20***
Task/Action	.31	.43	.72	−.36***
Results	.15	.22	.59	−.36***

* $p < .05$; ** $p < .01$; *** $p < .001$.

used to measure teamwork and organizing focus on more everyday situations, making it more difficult to find a specific situation to tell about (Bangerter et al., 2014). On the contrary, the questions asked to measure adaptation and autonomy focus more on exceptional situations.

Informing applicants did not help them produce more stories. This may be due to several reasons. First, the applicants may not have had enough time to prepare. Second, the example given may not have helped them to prepare sufficiently for the competencies that were measured. Third, the example of a question and the instruction to tell what they did in a specific situation may not have given them sufficiently explicit guidance for producing stories. In Study 3, we explored more explicit modes of providing information.

Study 3

Study 3 further tested H1 as well as H1b. We again conducted an experiment featuring a simulated selection interview with participants playing the role of applicants. We manipulated the *information* variable in three conditions. The first was the same *uninformed* condition as in Study 2, where applicants received no information at all about upcoming questions. In the second condition, *informed about competencies*, applicants were informed about the four competencies questions would focus on. Informing applicants about the competencies measured allows them to focus their search for a relevant example rather than wasting resources on guessing criteria (Klehe, König, Richter, Kleinmann, & Melchers, 2008; Kleinmann et al., 2011). Applicants may thus have more resources available to search their memory for a relevant episode for each specific competence. Finally, we developed a third *fully informed* condition, where applicants received information about the four competencies (as in the *informed about competencies* condition), a definition of a past-behaviour question with an example and watched a video example of a fictional applicant narrating a story to answer a past-behaviour question. Following the failure of the information manipulation in Study 2, this condition tested whether modelling an answer (Taylor, Russ-Eft, & Chan, 2005) would help applicants improve their storytelling performance. To test HI and H1b (*information decreases production of pseudostories*), we coded the presence or absence of both stories and pseudostories in applicants' responses.

Study 3 also investigated RQ2 (*What are the cognitive and emotional processes associated with answering past-behaviour questions?*). We adopted a method from Hall and Schmid Mast (2007): After participating in the simulated interview, applicants viewed the videorecording of how they answered questions during the interview. They were asked to write down all the thoughts and emotions they experienced during their performance.

Finally, Study 3 investigated RQ3, the effect of competency on the story production.

Methods

Participants

We recruited 83 students (41 women, *Mage* = 22.02, *SDage* = 2.16) to participate. A majority (65) had previously held a student job and had previously experienced at least one selection interview (69) (*M* = 3.20 interviews, *SD* = 3.19). Participants were randomly assigned to one of the three conditions: *uninformed* condition (26 participants), *informed about competencies* condition (27 participants) and *fully informed* condition (30 participants). Participants each received 15 Swiss francs for participating and three additional prizes of 50 Swiss francs were awarded for the three best performances during the interview (one per condition).

Procedure

Procedure was similar to Study 2, with minor differences. Experimenters were the first author and participants in a seminar on personnel selection. They were trained to conduct the interview in a standardized manner following a strict protocol. Each experimenter recruited between three and five participants (excluding psychology students and personal contacts) and conducted the interview with them.

Competencies measured were communication, autonomy, organizing and adaptation. Participants had 10 minutes before the interview to prepare. In the *uninformed* condition, they received these instructions: *You have 10 minutes to prepare yourself. By the end of the preparation, the interview will begin. You will have to introduce yourself. Then, you will have to respond to job-related questions. Your goal is to be hired, be convincing and perform!* In the *informed about competencies* condition, they received the same instructions as in the *uninformed* condition as well as additional information about competencies: *Here are the four competencies you will be asked about: communication, autonomy, organization and adaptation*. In the *fully informed* condition, they received the same instructions as in the *informed about competencies* condition as well as additional information about past-behaviour questions (*You will be asked past-behaviour questions. This type of question refers to a situation, a behaviour you have shown in the past. Here is an example of past-behaviour question about the competency of persuasion: "Could you tell me about a situation where you managed to convince a person about a thing, an opinion or to do something she would never have done without your intervention, in your private, professional or student life?". The expected*

response is a story: you have to tell a specific and concrete episode which demonstrates that you master the competence. You can now play the example of a response) and then watched an example video of someone producing a story to respond to a past-behavir question.

The interview began with an introductory question (*Could you present yourself and summarize your career in a few words?*) in order to increase the realism of the simulation. Then, participants were asked four past-behaviour questions. The experimenter did not use any probes, except in two specific situations: when participants did not say anything for several seconds (*Take your time to think*) or when they talked about a generic situation that did not involve storytelling (*Could you give me an example of a specific situation)*. The whole interview was videotaped. After the interview, participants filled out a personality survey (NEO PI-R). Then, they were invited to watch the videotape of their performance during the interview and write down any thought or feeling they remembered having experienced during the interview, as well as the time on the video. Participants were instructed not to report thoughts or feelings they experienced while watching the video (Hall & Schmid Mast, 2007). They then completed the Wonderlic Personnel Test and an ATIC measure. Finally, participants were debriefed and paid. Winners of the prizes were contacted after the study by email. Interviews were transcribed as in Study 2.

Measures

Interview experience, GMA, ATIC (Cohen's kappa = .72) and personality (neuroticism alpha = .69, extraversion alpha = .78, openness alpha = .74, agreeableness alpha = .61 with one item removed and conscientiousness alpha = .76) measures were identical to those used in Study 2 (descriptives appear in Table 4).

Stories. Coding of stories was the same as in Study 2 (double-coding 10% of the dataset, Cohen's kappa = .67).

Pseudostories. A pseudostory was defined as "a description of a generic situation or recurrent set of similar situations, without unity of time or action. It differs from a story in that it is a description of several events rather than of a unique event" (Bangerter et al., 2014, p. 6). We coded the presence or absence of pseudostories for the responses of each past-behaviour questions (double-coding of 10% of the dataset, Cohen's kappa = .67). We computed a total score of pseudostories (0–4) by participant over all four questions.

Comments. After the interview, participants viewed the video of their question-answering and noted their comments. All comments were segmented into single utterances (with one subject, verb and object). We content-analysed each utterance, attributing it to one of 16 mutually exclusive categories (Cohen's kappa = .87 based on double-coding of 130 comments).

Results

We found significant correlations between conscientiousness, interview experience and the production of stories, but no significant correlations between personality, interview experience and the production of pseudostories. We found no significant correlations between age, gender, ATIC and the production of stories or pseudostories (see Table 4). There were also some significant correlations between some of these variables and the comments participants made, which we will discuss below.

Hypothesis 1: Effects of information on storytelling

To test H1, we computed two hierarchical regression models. In Model 1, given that story production was significantly positively correlated with conscientiousness and interview experience, we entered these variables as predictors of the proportion of responses featuring stories (Table 5). Both conscientiousness and interview experience emerged as positive significant predictors and the model explained a significant proportion of variance. In Model 2, we coded the three information conditions into dummy variables (*informed about competencies*

Table 4. Means (*M*), standard deviations (*SD*) and correlations for main study variables (Study 3).

	Variable	M	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	Condition	1.04	0.83																		
2	Gender	0.49	0.50	.07																	
3	Age	22.02	2.16	.05	.01																
4	Interview experience	3.20	3.19	−.09	−.01	.26*															
5	ATIC	2.34	1.08	.04	.00	−.16	−.01														
6	Extraversion	4.17	0.81	.14	.20	−.10	.03	.13													
7	Agreeableness	4.66	0.63	−.04	.08	.08	−.04	−.07	−.25*												
8	Conscientiousness	4.77	0.72	−.11	.26*	−.19	−.05	.07	−.05	.39**											
9	Neuroticism	2.75	0.65	.09	.07	.04	−.13	−.12	−.01	−.11	−.32**										
10	Openness	4.02	0.78	−.01	.01	.12	.21	−.12	.04	.01	−.10	−.10									
11	GMA	24.00	5.39	−.03	−.20	−.16	−.04	.20	.01	.05	−.08	−.10	.01								
12	Stories	2.10	1.28	.11	.08	.10	.27*	−.04	.09	.14	.27*	−.19	.19	−.07							
13	Pseudostories	2.18	1.27	−.24*	−.12	−.07	−.15	.11	−.09	−.03	−.15	.05	−.09	.14	−.80**						
14	Factor 1: Resp. prod.	0.00	1.00	−.13	.06	.03	.00	.05	.00	−.05	−.01	.09	.01	.00	−.16	.22					
15	Factor 2: Find. ex.	0.00	1.00	.10	.16	−.01	−.11	.11	.18	−.15	.18	−.10	−.17	.08	.05	.11	.00				
16	Factor 3: Non-verb.	0.00	1.00	−.08	.04	.21	.09	.13	.07	.12	.07	−.16	.10	.13	.02	.05	.00	.00			
17	Factor 4: Neg. emot.	0.00	1.00	.12	.12	.07	−.03	.06	−.02	.02	.02	.26*	.05	.10	−.05	−.06	.00	.00	.00		
18	Factor 5: Pos. emot.	0.00	1.00	−.08	−.20	.07	.04	.25*	−.11	−.08	−.05	.07	−.02	.20	−.04	.06	.00	.00	.00	.00	
19	Factor 6: Faking	0.00	1.00	.08	−.06	.10	.06	.17	−.19	.10	.08	−.12	.07	.05	.18	−.09	.00	.00	.00	.00	.00

Condition: 0 = uninformed, 1 = informed about competencies, 2 = fully informed. Gender: 0 = Men, 1 = Women. ATIC: Ability to identify criteria. GMA: General mental ability. Factor 1: Resp. prod.: response production in general. Factor 2: Find. ex.: Finding an example, trouble finding a relevant example. Factor 3: Non-verb.: Non-verbal behaviour and time management. Factor 4: Neg. emot.: Negative emotional state, fear of finding no response. Factor 5: Pos. emot.: Positive emotional state, no problem producing a response.

* $p < .05$; ** $p < .01$; *** $p < .001$.

Table 5. Summary of hierarchical multiple regression predicting stories.

	Model 1		Model 2	
Variable	B	SE B	B	SE B
Intercept	−.73	.91	−1.53	.96
Interview experience	.12**	.04	.12**	.04
Conscientiousness	.52**	.19	.60**	.19
Informed about competencies			.69*	.33
Fully informed			.54	.32
Adj. R^2		.13		.16
F for R^2 change		7.32**		2.45

$N = 83$. * $p < .05$; ** $p < .01$; *** $p < .001$.

condition and *fully informed* condition) and used them as additional predictors. These variables did not explain a significant additional proportion of variance of the production of stories (Table 5), meaning that information in general did not influence the participants' production of stories. Therefore, H1 was not supported. Nevertheless, the *informed about competencies* condition appeared as a significant positive predictor.

Hypothesis 1b: Effects of information on pseudostories

To test H1b, we computed the same regression models as for H1 but predicting the proportion of responses featuring pseudostories. Results of Model 1 showed that conscientiousness and interview experience did not predict a significant proportion of variance of the production of pseudostories. Results of Model 2 showed that information predicts a significant proportion of additional variance of the production of pseudostories (Table 6). Conscientiousness and both the *informed about competencies* condition and the *fully informed* condition emerged as significant negative predictors of pseudostory production, thus supporting H1b.

Research question 2: cognitive and emotional processes

We coded 825 comments at the utterance level into 16 categories (Table 7). Comments regarding finding an example and negative emotional states are the most frequent, making up 23% of comments and being mentioned by a majority of participants.

We conducted an exploratory factor analysis (varimax rotation) on the categories to determine their factor structure. There were 6 factors with eigenvalue >1, explaining 63.82% of variance. Table 8 shows the factors and the descriptive labels we gave them. The first two factors, *response production in general* and *finding an example*, explain 31% of variance. Factor 4 (*negative emotions*) correlates with neuroticism, suggesting that this personality trait leads participants to experience more negative emotions when responding to past-behaviour questions. Factor 5 (*positive emotions*) correlates with the ability to identify criteria (ATIC). This correlation is difficult to interpret, but suggests that participants able to identify criteria might possibly feel less stressed and more relieved because they are better able to identify the competencies and thus potentially better able to respond.

Table 6. Summary of hierarchical multiple regression predicting pseudostories.

	Model 1		Model 2	
Variable	B	SE B	B	SE B
Intercept	3.69***	.95	4.85***	.98
Interview experience	−.06	.04	−.07	.04
Conscientiousness	−.28	.19	−.39*	.19
Informed about competencies			−.95**	.34
Fully informed			−.85*	.32
Adj. R^2		.02		.11
F for R^2 change		1.92		4.96**

$N = 83$. * $p < .05$; ** $p < .01$; *** $p < .001$.

Research question 3: storytelling by competency

Using Bonferroni correction ($p = .05/6 = .008$), we found significant differences in story production by competency. Organizing ($M = .34$, $SD = .48$) was significantly different from communication ($M = .54$, $SD = .50$), $t(82) = 3.11$, $d = 0.68$, $p = .003$, from autonomy ($M = .61$, $SD = .49$), $t(82) = 4.00$, $p < .001$, $d = 0.88$, and from adaptation ($M = .60$, $SD = .49$), $t(82) = 3.86$, $p < .001$, $d = 0.85$. Communication was not significantly different from autonomy, $t(82) = 1.14$, $p = .26$, $d = 0.25$, and from adaptation, $t(82) = .96$, $p = .34$, $d = 0.21$. Autonomy was not significantly different from adaptation, $t(82) = .18$, $p = .86$, $d = 0.04$. Participants thus produced significantly less stories when asked about organizing.

Discussion

Regarding Hypothesis 1, information did not significantly influence the production of stories, because the overall regression model did not explain sufficient variance. Regarding Hypothesis 1b, information significantly decreased the production of pseudostories (both conditions were effective relative to the control condition). Thus, providing applicants with information about what is expected of them does not reliably increase story production, but it does decrease inappropriate responding (pseudostories). This suggests that providing information increases awareness about what constitutes an (in)appropriate answer to a past-behaviour question, without actually helping applicants to produce an appropriate answer.

Regarding Research Question 2, results suggest that the most frequent obstacle to applicants' production of stories is trouble immediately finding a relevant example to recount. This may perhaps play a more important role in applicants' responses than a lack of understanding about what is expected of them, because even in the *fully informed* condition, such comments were frequent. A good illustration is the following comment made by one applicant: "*Shit! All questions are going to be in this form. I'm going to have to tell stories. I anticipated more general questions like 'Why are you a good candidate' […] It's really going to be just questions like that. I think I'm a good narrator. I just have to find anecdotes*". The fine-grained categories in Table 7 suggest that many aspects of responding are difficult for applicants, especially response quality (e.g., the level of detail to use, 38.6% of applicants) or discourse production (e.g., having trouble to find one's words, 49.4% of applicants). These results converge with the significant effect of the information-providing manipulations on the production of pseudostories (applicants are more aware that stories are appropriate and pseudostories are inappropriate) and the lack of significant effects of information on the production of stories (they need more help to actually find an example and put it into words).

The second most frequently mentioned category in Table 7 relates to negative emotional reactions, like feelings of surprise,

Table 7. Categories of comments (Study 3).

Category		Definition/Content of the comment	Frequency	Mention (%)
1	Finding an example	Trouble experienced in finding an example, hesitation between two examples (the word "example" should be explicitly mentioned)	98	60.2
2	Negative emotional state	Negative emotional state, for example stress, hesitation, disappointment, uncertainty, surprise, apprehension	95	56.6
3	Other	Irrelevant comments, for example *"It reminds me of good memories"*	76	47
4	Positive emotional state	Positive emotional state, for example joy, relief, satisfaction	75	48.2
5	Response quality	Reflection on the quality of the response, the vocabulary used, the level of detail	73	38.6
6	Discourse production	Trouble finding words, how to speak	69	49.4
7	Irrelevance of the example	Example judged too general, irrelevant, or narrated just to say something (the word "example" should be explicitly mentioned)	60	39.8
8	Finding no response	Fear of not finding a response to the question asked, having no idea of response, not knowing what to respond	43	32.5
9	Questions asked	Reflection on questions asked, on job requirements or on information previously read in the job advertisement	41	30.1
10	Relevance of the example	Reflection on the relevance of the example told, judged as a good one, easily found (the word "example" should be explicitly mentioned)	38	34.9
11	Example used repeatedly	Reflection on the repeated use of the same example for different questions, need to find another example than the one already told	34	26.5
12	Type of question asked	Reflection on the type of question that was coming, apprehension about the question type	34	27.7
13	Faking	Inventing examples or embellishing the situation	29	24.1
14	Non-verbal behaviour	Participant's own non-verbal behaviour	26	15.7
15	Temporization	Temporizing by asking a question, by repeating the question, by giving useless information or by saying *"uh"*	19	16.9
16	Turn management	Worries about turn-taking, the pause before the response or the duration of the response	15	18.1

$N = 825$. Mention (%): % of participants mentioning the category at least once.

Table 8. Categories (after factor analysis) (Study 3).

Factor	Variance (%)	Description	Examples
1	20.31%	Response production in general	*I don't know if I really answered the question/I'm not sure I understand the question and the answer to tell*
2	10.70%	Finding an example, trouble finding a relevant example	*I tell myself there it is she wants a specific example but I don't have one/Difficulties of finding a specific example to tell*
3	9.56%	Non-verbal behaviour and time management	*I move my shoulders non-stop, a sign of tension I should avoid/I move too much/I needed time to think a little bit*
4	8.38%	Negative emotional state, fear of finding no response	*I apprehend the question/I don't know what to say/Feeling of insecurity/I'm completely lost*
5	7.95%	Positive emotional state, no problem producing a response	*I felt good and confident/It makes me feel proud in a sense I felt no hesitation*
6	6.91%	Faking	*Question to which I respond by improvising/I embellish the situation/I totally make up something*
Total	63.81%		

disappointment and stress. These emotions may be triggered by the unanticipated nature of past-behaviour questions (especially in the *uninformed* condition) and interact with the above-mentioned difficulties in finding examples to further inhibit applicants' responding. Thus, responding to past-behaviour questions may be associated with interview anxiety (McCarthy & Goffin, 2004).

Applicants' comments also suggested that phenomena like faking might be related to difficulties in finding examples. Several applicants admitted making up examples as a result of unsuccessful searches of their memory (e.g., *"No more inspiration – I make something up"* or *"I can't find an example – I make up an example"*). These findings are particularly interesting because models of the determinants of faking in interviews typically do not specify the local interactional situations in which faking occurs (Levashina & Campion, 2006). Here, applicants' comments suggest that faking may constitute an ad hoc response to a self-presentational dilemma. In general, then, these exploratory data suggest several avenues for further research on processes behind responding to past-behaviour questions.

Regarding Research Question 3, as in Study 2, there were differences regarding the four competencies when applicants were asked about communication, adaptation and autonomy, they produced more stories than when asked about organizing. Similar results were found in Study 2 and in previous research (Bangerter et al., 2014). Since the order in which the competencies were assessed is not the same as in Study 2, it seems likely that some competencies are more difficult to translate into stories than others, especially routine competencies like organizing.

Finally, interview experience and conscientiousness were both correlated with story production. Applicants with more interview experience may have been exposed to past-behaviour questions, or to recruiters' questions asking about examples to illustrate particular claims. These may have made them more aware of the expected answers or helped them find examples. That more conscientious applicants produce more stories may suggest a motivational effect. Conscientious applicants may have been more motivated to conform to expectations, and thus may have invested more effort in finding an example to narrate.

General discussion

The rise of behavioural interviewing will make storytelling skills more important. But initial research on applicants' responses to past-behaviour questions (Bangerter et al., 2014) suggests that applicants may have difficulties producing stories. Here, we first conducted a field study to ascertain the efficacy of past-behaviour questions in eliciting stories. Then, we conducted two experiments manipulating the effects of information and probing on the production of stories and pseudostories and we explored applicants' experiences in responding to past-behaviour questions, as well as whether storytelling was affected by the competency measured.

Our research has produced important results. Study 1 suggests that simply asking past-behaviour questions is not sufficient to elicit an appropriate response, even in a real, high-stakes selection situation. Indeed, 50% of the time, applicants failed to respond appropriately. Note that this rate is similar to story production in the experiments (56% in Study 2 and 53% in Study 3), suggesting that they do not really lack ecological validity in this respect.

Study 2 showed that brief information about the type of questions and how to respond did not significantly increase production or narrative diversity of stories. In Study 3, providing information about competencies and providing a concrete example of how to respond did not significantly increase story production. However, both kinds of information significantly reduced the production of pseudostories. This pattern shows that even explicit instructions do not help applicants produce more stories, even though they may reduce inappropriate responses like pseudostories. This converges with findings from applicants' reports of the cognitions and emotions they experienced in Study 3: Their difficulties seem to be recalling a suitable example and putting it into words, or recalling enough examples to answer multiple questions, rather than being unaware that stories are expected, or hesitating to take an extended turn.

The probing manipulation in Study 2 was successful in inducing applicants to produce more stories and to increase the narrative diversity of stories. This suggests that recruiters may have an important role to play in eliciting relevant applicant responses when using past-behaviour questions, thus increasing information accuracy and completeness (Levashina et al., 2014). These findings further support the idea of storytelling as a collaborative process (Bavelas et al., 2000) in the specific context of selection interviews.

Finally, we consistently found that story production varied as a function of the competency measured, suggesting that some competencies (that potentially reflect more routine, everyday behaviour) may be more difficult to translate into a story.

Theoretical implications

Our findings speak to at least two theoretical issues. First, they respond to calls for research on probing in behavioural interviews (De Kock et al., 2018; Levashina et al., 2014). Our findings may seem to contradict research on interview structure, which advocates limiting probing to reduce bias (Campion et al., 1997; Dipboye, Wooten, & Halverson, 2004; Huffcutt & Arthur, 1994). However, the effects of increasing structure on validity are subject to a ceiling (Huffcutt & Arthur, 1994). In a four-level scale of structure, levels 3 and 4 do not differ in terms of validity (respectively, .56 and .57). And "level 3 was characterized by prespecification of the questions, although applicants were not asked the exact same questions because of the use of different interview forms or allowing interviewers to choose among alternative questions and to probe responses to the specified questions" (Huffcutt & Arthur, 1994, p. 186). Thus, probing may be considered part of the leeway recruiters may retain during the interview without endangering validity. Furthermore, we used planned probing, a specific type of probing that has been hypothesized to "help applicants who might be shy or speak in succinct ways to clarify their answers and provide more detailed job-related information" or "help applicants recall and identify the most relevant experiences they need to describe in order to best present their true job-related credentials" (Levashina et al., 2014, p. 272). Although this has not been tested, planned probing might even increase validity and lead to more positive user reactions (Levashina et al., 2014). Probing also raises questions regarding faking. On the one hand, it may increase faking, especially in response to past-behaviour questions with inexperienced applicants. But it may also reduce opportunities to fake, because "with no follow-up, respondents might think that there is no way that the interviewer can detect faking" (Levashina & Campion, 2006, p. 309). Indeed, interviewers may use probes to flag incomplete or inconsistent information and potentially alert applicants about their abilities to detect lying. Overall, then, our findings bring a much-needed data point to the study of probing in behavioural interviews, showing that it can increase the completeness or relevance of applicants' responses by encouraging the production and content of stories.

Beyond behavioural interviewing, our findings inform the general problem of elicitation of trait-expressive behaviour in personnel selection. According to trait activation theory, expression of trait-relevant behaviours is moderated by the situation. Lievens et al. (2015) distinguish between general and specific levels of eliciting trait-relevant behaviours in selection situations, with a focus on assessment centres. At a general level, how an assessment centre exercise is designed (including its instructions) can serve as an elicitor. At a specific level, elicitors can be embedded in the situation, for example, when role-players are instructed to display specific emotional reactions. Schollaert and Lievens (2011) showed that instructing actors to use prompts leads to better measurement of some dimensions without discernible negative impact. These findings are analogous to ours: Providing information about expected responses (a general-level elicitor) was less effective in eliciting stories than allowing recruiters to probe (a specific-level elicitor). Just as "good raters" (De Kock et al., 2018) actively contribute to expression of relevant cues through interactions with applicants, recruiters may enhance opportunities to construct better measurements via probing.

Limitations

Our research has limitations. First, Studies 2 and 3 were experimental simulations of selection interviews, with nothing major at stake for participants except the additional prizes for the best performances. Despite the fact that experimental participants

often are highly motivated and treat simulated interviews like real ones (see, for example, Swider, Barrick, Harris, & Stoverink, 2011), generalizability of our findings to real interviews remains unknown. However, the rate of storytelling from the field study (Study 1) was very similar, thus suggesting that motivational differences may not have impacted the phenomena under study. Second, Study 2 and 3 participants were students with little professional experience. Participants with more experience may have had more anecdotes to recount and may therefore have reacted differently to the manipulations. Third, the study sample size was low, which may have decreased chances of finding effects (especially for information). Fourth, we did not counterbalance the order in which competencies were measured. Therefore, the finding that some competencies are more amenable to the production of narrative responses remains tentative (however, those findings do corroborate previous research; Bangerter et al., 2014).

Practical implications

Our findings have straightforward practical implications. First, recruiters cannot assume that applicants will produce optimal responses to past-behaviour questions without some form of encouragement. Nor can they assume that brief instructions will be effective. Recruiters may be interested in the potential of probing to enhance storytelling responses. However, the planned probes we used are different from the freewheeling interactional style many recruiters prefer (Lievens & De Paepe, 2004). Training programmes for recruiters should thus emphasize planned probing in the context of structured interview procedures. Finally, because behavioural interviews often measure multiple competencies, recruiters should be aware that not all competencies are equally easy to translate into a story, and that applicants may have difficulty finding multiple examples. Our findings also have implications for preparing applicants for behavioural interviews. Given the negligible impact of information on storytelling production, it may be necessary to apply more in-depth coaching interventions (Ralston et al., 2003).

Future research

Future research might explore the difficulties of producing stories by testing whether giving applicants more time to prepare for the interview increases the production of narrative responses by helping them find an example or choose the best example among those they may have available. This possibility is supported by recent findings that interviewees who have more time to think about what to recount end up talking about different experiences that may better reflect their maximal performance level (Huffcutt et al., 2016).

Given the lack of impact of information provision on storytelling, the efficacy of more intensive, coaching-based interventions to increase storytelling might be evaluated. These would entail inviting applicants to think about their particular competencies as well as about occasions when they demonstrated mastery of those competencies, before reflecting on how to translate those occasions into informative stories (Ralston et al., 2003).

More generally, we advocate for more research on factors affecting narrative responding to past-behaviour questions. One line of research might seek to specify the criteria for effective storytelling, i.e., looking at storytelling ability as orthogonal to specific competencies past-behaviour questions measure. To guarantee construct validity, it is important to distinguish what facets of a response reflect general storytelling ability versus mastery of a specific competency. While general criteria for storytelling quality exist (Ralston et al., 2003), there is no research empirically testing their validity. This is important because storytelling ability is an important facet of personal abilities like leadership (Sharma & Grant, 2011). Our measure of storytelling quality is crude: We measure its presence or absence and the number of basic narrative elements. But more qualitative aspects could be measured, e.g., how captivating or convincing a story is. Such measures might build on the Ralston et al. (2003) criteria and might potentially uncover links to abilities like leadership.

Finally, more research on the effects of probing is necessary (De Kock et al., 2018; Levashina et al., 2014). This need is particularly urgent for the effects of probing on faking because existing results are inconsistent. But the impact of probing also needs to be evaluated in terms of interrater reliability and predictive validity.

Conclusions

The behavioural interview is a selection method with desirable psychometric properties like validity, reliability, and robustness against faking. But little is known regarding its properties as a social interaction between recruiters and applicants. Our research demonstrated difficulties involved in producing stories in response to past-behaviour questions, as well as suggesting ways recruiters may actively enable applicants to produce representative, meaningful stories. We hope to stimulate more research on this important topic.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung [100019_152920].

References

Bangerter, A., Corvalan, P., & Cavin, C. (2014). Storytelling in the selection interview? How applicants respond to past behavior questions. *Journal of Business and Psychology*, *29*, 593–604.

Bangerter, A., & Gosteli-Corvalan, P. (2017). Taking the initiative in job interviews: Extended responses to questions and storytelling. In S. Pekarek Doehler, A. Bangerter, G. de Weck, L. Filliettaz, E. González-Martínez, & C. Petitjean (Eds.), *Interactional competences in institutional settings* (pp. 117–142). Cham: Palgrave Macmillan.

Bangerter, A., Roulin, N., & König, C. J. (2012). Personnel selection as a signaling game. *Journal of Applied Psychology*, *97*, 719–738.

Bavelas, J. B., Coates, L., & Johnson, T. (2000). Listeners as co-narrators. *Journal of Personality and Social Psychology*, *79*, 941–952.

Bradburn, N. M., Rips, L. J., & Shevell, S. K. (1987). Answering autobiographical questions: The impact of memory and inference on surveys. *Science*, *236*, 157–161.

Brosy, J., Bangerter, A., & Mayor, E. (2016). Disfluent responses to job interview questions and what they entail. *Discourse Processes*, *53*, 371–391.

Campion, M. A., Palmer, D. K., & Campion, J. E. (1997). A review of structure in the selection interview. *Personnel Psychology*, *50*, 655–702.

Costa, P. T., & McCrae, R. R. (1992). Normal personality assessment in clinical practice: The NEO personality inventory. *Psychological Assessment*, *4*, 5–13.

De Kock, F., Lievens, F., & Born, M. (2018). The profile of the 'Good Judge'in HRM: A systematic review and agenda for future research. *Human Resource Management Review*, 100667.

Dipboye, R. L. (1994). Structured and unstructured selection interviews: Beyond the job-fit model. In G. Ferris (Ed.), *Research in personnel and human resources management* (Vol. 12, pp. 79–123). Greenwich: JAI Press.

Dipboye, R. L., Wooten, K., & Halverson, S. (2004). Behavioral and situational interviews. In J. Thomas (Ed.), *Comprehensive handbook of psychological assessment, industrial and organizational assessment* (Vol. 4, pp. 297–316). Hoboken, NJ: Wiley.

Ellis, A. P., West, B. J., Ryan, A. M., & DeShon, R. P. (2002). The use of impression management tactics in structured interviews: A function of question type? *Journal of Applied Psychology*, *87*, 1200–1208.

Hall, J. A., & Schmid Mast, M. (2007). Sources of accuracy in the empathic accuracy paradigm. *Emotion*, *7*, 438–446.

Huffcutt, A. I., & Arthur, W. (1994). Hunter and Hunter (1984) revisited: Interview validity for entry-level jobs. *Journal of Applied Psychology*, *79*, 184–190.

Huffcutt, A. I., Conway, J. M., Roth, P. L., & Klehe, U. C. (2004). The impact of job complexity and study design on situational and behavior description interview validity. *International Journal of Selection and Assessment*, *12*, 262–273.

Huffcutt, A. I., Culbertson, S. S., Goebl, A. P., & Toidze, I. (2016). The influence of cognitive ability on interviewee performance in traditional versus relaxed behavior description interview formats. *European Management Journal*, *35*, 383–387.

Huffcutt, A. I., Weekley, J. A., Wiesner, W. H., Degroot, T. G., & Jones, C. (2001). Comparison of situational and behavior description interview questions for higher-level positions. *Personnel Psychology*, *54*, 619–644.

Judge, T. A., Cable, D. M., & Higgins, C. A. (2000). The employment interview: A review of recent research and recommendations for future research. *Human Resource Management Review*, *10*, 383–406.

Kessler, R. (2006). *Competency-based interviews*. Franklin Lakes, NJ: Career Press.

Klehe, U.-C., König, C. J., Richter, G. M., Kleinmann, M., & Melchers, K. G. (2008). Transparency in structured interviews: Consequences for construct and criterion-related validity. *Human Performance*, *21*, 107–137.

Kleinmann, M., Ingold, P. V., Lievens, F., Jansen, A., Melchers, K. G., & König, C. J. (2011). A different look at why selection procedures work The role of candidates' ability to identify criteria. *Organizational Psychology Review*, *1*, 128–146.

Labov, W. (1972). *Language in the inner city: Studies in the Black English vernacular* (Vol. 3). Philadelphia: University of Pennsylvania Press.

Levashina, J., & Campion, M. A. (2006). A model of faking likelihood in the employment interview. *International Journal of Selection and Assessment*, *14*, 299–316.

Levashina, J., Hartwell, C. J., Morgeson, F. P., & Campion, M. A. (2014). The structured employment interview: Narrative and quantitative review of the research literature. *Personnel Psychology*, *67*, 241–293.

Lievens, F., & De Paepe, A. (2004). An empirical investigation of interviewer-related factors that discourage the use of high structure interviews. *Journal of Organizational Behavior*, *25*, 29–46.

Lievens, F., Schollaert, E., & Keen, G. (2015). The interplay of elicitation and evaluation of trait-expressive behavior: Evidence in assessment center exercises. *Journal of Applied Psychology*, *100*, 1169–1188.

Lonati, S., Quiroga, B. F., Zehnder, C., & Antonakis, J. (2018). On doing relevant and rigorous experiments: Review and recommendations. *Journal of Operations Management*, *64*, 19–40.

Mandelbaum, J. (2013). Storytelling in conversation. In J. Sidnell & T. Stivers (Eds.), *The handbook of conversation analysis* (pp. 492–508). Chichester: John Wiley and Sons.

McCarthy, J., & Goffin, R. (2004). Measuring job interview anxiety: Beyond weak knees and sweaty palms. *Personnel Psychology*, *57*, 607–637.

Norrick, N. R. (2007). Conversational storytelling. In D. Herman (Ed.), *The Cambridge companion to narrative* (pp. 127–141). Cambridge: Cambridge University Press.

Podsakoff, P. M., & Podsakoff, N. P. (in press). Experimental designs in management and leadership research: Strengths, limitations, and recommendations for improving publishability. *The Leadership Quarterly, 30*, 11–33.

Ralston, S. M., Kirkwood, W. G., & Burant, P. A. (2003). Helping interviewees tell their stories. *Business Communication Quarterly*, *66*, 8–22.

Roulin, N., & Bangerter, A. (2012). Understanding the academic–Practitioner gap for structured interviews: 'Behavioral'interviews diffuse,'structured'interviews do not. *International Journal of Selection and Assessment*, *20*, 149–158.

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

Schank, R. C., & Berman, T. R. (2002). The pervasive role of stories in knowledge and action. In M. C. Green, J. J. Strange, & T. C. Brock (Eds.), *Narrative impact: Social and cognitive foundations* (pp. 287–313). Mahwah, NJ: Lawrence Erlbaum.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*, 262–274.

Schollaert, E., & Lievens, F. (2011). The use of role-player prompts in assessment center exercises. *International Journal of Selection and Assessment*, *19*, 190–197.

Sharma, A., & Grant, D. (2011). Narrative, drama and charismatic leadership: The case of Apple's Steve Jobs. *Leadership*, *7*, 3–26.

Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language*, *32*, 25–38.

Stevens, C. K., & Kristof, A. L. (1995). Making the right impression: A field study of applicant impression management during job interviews. *Journal of Applied Psychology*, *80*, 587–606.

Swider, B. W., Barrick, M. R., Harris, T. B., & Stoverink, A. C. (2011). Managing and creating an image in the interview: The role of interviewee initial impressions. *Journal of Applied Psychology*, *96*, 1275–1288.

Taylor, P. J., Russ-Eft, D. F., & Chan, D. W. (2005). A meta-analytic review of behavior modeling training. *Journal of Applied Psychology*, *90*, 692–709.

Van Iddekinge, C. H., McFarland, L. A., & Raymark, P. H. (2007). Antecedents of impression management use and effectiveness in a structured interview. *Journal of Management*, *33*, 752–773.

Wonderlic, E. G. (2001). *Wonderlic personnel test manual*. Northfield, IL: E. F. Wonderlic & Assoc.