Methodological Challenges and Advances in Managerial and Organizational Cognition
New Horizons in Managerial and Organizational Cognition, 2018, 123–146
Edited by Robert J. Galavan, Kristian J. Sund and Gerard P. Hodgkinson
doi:10.1108/S2397-52102017006

CHAPTER

6

To Grasp Cognition in Action, Combine Behavioral Experiments with Protocol Analysis

Charlotte Reypens and Sheen S. Levine

ABSTRACT

With behavioral experiments and protocol analysis, researchers can capture cognition in action. Using behavioral -experiments, they can study realized behavior, not -perception or self-reports. And they can do that in a controlled laboratory environment to establish causality, curbing spurious -relationships. With protocol analysis, a method to elicit decision-makers' thoughts, researchers can tap into cognitive processes. In combination, the two methods offer a novel approach to grasp mental processes alongside behavior, to reach causality and replicate findings. We describe the methods, demonstrate how researchers can apply them, and share practices from the design of experimental instruments to the replication of findings.

Keywords: Experiment; protocol analysis; cognition

"The greatest challenge is one of measurement," concluded -Hodgkinson and Healey (2008, p. 387) after reviewing the research on cognition in organizations. Understanding managerial cognition may be paramount (Hodgkinson & Sparrow, 2002), but researchers must contend with methodological challenges when attempting to identify, isolate, measure, and grasp mental processes related to organizations.

To address those challenges, we suggest grasping cognition in an environment that simulates managerial decision-making, but with controls that facilitate clarity and allow determining causality. We present two methods: behavioral experiments and protocol analysis, the latter being a technique to elicit thought processes and use them as data (Ericsson & Simon, 1984).[1] By combining both methods, researchers can capture cognition in action: They can observe realized behavior while simultaneously documenting cognitive processes. And, because data are collected in a controlled environment, researchers can untangle cause and effect to identify causality and offer predictions—the ultimate goal of science (Popper, 1959, 1963). The simple setup also eases replication, a much-needed feature when "the truth is under attack," as Levine (2012) alerted (also see Bettis, Helfat, & Shaver, 2016; Lewin et al., 2016).

Cognition refers to "processes of knowing, including attending, remembering, and reasoning; also the content of the processes, such as concepts and memories" (American Psychological Association, 2009; also see Helfat & Peteraf, 2015; Levine, Bernard, & Nagel, 2017). In most human settings, let alone organizational ones with all of their complexity, a multitude of variables affect behavior and decision-making. So, researchers who wish to understand cognition need to identify relevant variables, examine relationships between them, and finally—pinpoint the cognitive processes underlying observed outcomes. Yet, students of management and organizational cognition face at least three methodological challenges when attempting to do so: capturing underlying cognitive processes, establishing causality, and ensuring validity.

Nowadays, analytical techniques are aplenty: from ANOVA and regression in its many flavors to data mining (Bruce & Kristine, 2008, p. 15) and content analysis, typically of managerial documents such as letters to shareholders (e.g., Nadkarni & Chen, 2014). All can identify correlations between variables—but they are necessarily silent on the underlying cognitive processes, because such processes cannot be pinpointed by correlational methods, however statistically sophisticated. Thus, correlational methods should be complemented with techniques that can capture the underlying processes, such as experiments and protocol analysis.

Even more than the sheer number of variables involved, the interactions between variables frustrate researchers attempting to understand complex relationships (Sommer & Loch, 2004). A multitude of individual and context factors are intertwined, blurring the link between these factors and the outcomes of decisions and outcomes. Thus, even after relevant variables have been identified, a researcher must disentangle links between them and distinguish mere correlation from causality. As Colquitt (2008, p. 616) acknowledges, "inferring causality is one of the most difficult aspects of scientific research," so he proceeds to recommend, in an editorial published in the *Academy of Management Journal*, to study behavior in a controlled setting—a laboratory. There, mechanisms can be tested, outcomes isolated, and alternative explanations controlled, so that causality can be determined (Cook, Campbell, & Day, 1979; Mill, 1883).

A researcher may succeed in identifying variables and determine the relationship between them, but how can she ascertain that her findings are valid? In recent years, we have been reminded how research often delivers findings that are irreproducible, and possibly false (Butler, Delaney, & Spoelstra, 2017; Chang & Li, 2015; Tsang & Yamanoi, 2016). Researchers have long warned that a large share of research findings (or even most; Ioannidis, 2005) is false, and recent gigantic replication projects have confirmed this suspicion, for example, in economics (Camerer et al., 2016), psychology (Open Science Collaboration, 2015), and cancer research (Begley & Ellis, 2012). In management, Goldfarb and King (2016, p. 168) estimate that two in five research results, or 40%, are not replicable. Bergh, Sharp, Aguinis, and Li (2017) report worse. Findings that cannot be replicated may not be true, so replication is essential to the scientific effort, now featured in the editorial policies of leading scholarly journals (Bettis et al., 2016; Desai, 2013; Eich, 2014). Arie Lewin, Founder of *Organization Science* and *The Journal of International Business Studies*, took a step further: In a recent editorial, he and a group of editors advocate recognition for researchers who make their instruments and data publicly available and for those who register their hypotheses and analysis plan before embarking on a study (Lewin et al., 2016). Yet, if replication of a study is prohibitively difficult or expensive, it is unlikely to be pursued, so findings can never be substantiated (Agarwal, Croson, & Mahoney, 2010). A method that can be easily and cheaply replicated is a method that can help science advance.

To address these challenges, we propose, researchers of managerial and organizational cognition will benefit from a blend of behavioral experiments and protocol analysis. This mix of methods can help researchers observe cognition alongside behavior, in a way that is replicable. Behavioral experiments enable the study of realized behavior in a controlled environment, supporting the causality and replicability of research findings. When combined with protocol analysis, researchers can simultaneously tap into various cognitive processes that underlie observed behavior. In the next section, we describe both methods and suggest how they can help researchers identify underlying processes, establish causality, and test for replicability. We end by illustrating how researchers can apply both methods, using examples from our own work.

How Experiments and Protocol Analysis can Benefit Cognition Research

PROTOCOL ANALYSIS: THE USE OF VERBAL REPORTS AS DATA

K. Anders Ericsson, an expert on expertise, and Herbert Simon, an expert on almost everything, developed a method to elicit participants' thought processes and use their verbal reports as data—protocol analysis. Although researchers have been collecting verbal protocols since the early twentieth century, Ericsson and Simon (1984, 1993) provide a much needed guide to reliably use such protocols as data. Before that, researchers used protocols to study phenomena such as problem-solving (Simmel, 1953) or children's development (Hanfmann & Kasanin, 1937), often assuming that protocols automatically reflect cognitive processes. In *Protocol Analysis: Verbal Reports as Data*, Ericsson and Simon draw on theories of cognitive and interpretive processes to guide how verbal reports can be collected and interpreted systematically. When these guidelines are followed, "thinking aloud" does not interfere with participants' ability to think or their performance, according to the empirical evidence reviewed by Ericsson and Simon (1984, 1993). In a recent meta-analysis of 94 studies that adopted the technique, Fox, Ericsson, and Best (2011) confirm this observation.

Protocol analysis has applications in psychology, for example in understanding motivational processes (Robie, Brown, & Beaty, 2007); and in marketing, as in studying the usability of a product or service (Li, Daugherty, & Biocca, 2001). In management, Payne (1976) used the technique in studying information processing under varying levels of task complexity, examining how people search for alternatives, evaluate them, and choose an option. In another application of protocol analysis, Isenberg (1986) asked general managers and undergraduates to verbalize their thoughts while solving a business case, allowing him to compare problem-solving between both groups. Likewise, Lee and Puranam (2015) used the technique to contrast experts and novices in organizational design. Clark, Li, and Shepherd (2017) collected verbal protocols from managers to examine foreign market selection. Recently, protocol analysis appeared in entrepreneurship research: Dew, Read, Sarasvathy, and Wiltbank (2009) used it to compare how expert and novice entrepreneurs make decisions. Baron and Ensley (2006) and Grégoire, Barr, and Shepherd (2010) used it to study the reasoning processes of entrepreneurs as they recognize opportunities (or miss them). We also recognize an opportunity: Researchers who rely on protocol analysis can identify and interpret a range of cognitive processes.

To collect protocols, participants verbalize their thoughts as they are performing a task. They are asked to “think out loud” while focusing on the task at hand. Participants are not asked to explain their thoughts—just to verbalize them as they emerge. During the verbalizing process, participants search their minds to report what they are thinking (Gould, 1999). Verbal reports can be collected during the task (concurrently), after (retrospectively), or both. When protocol analysis is conducted concurrently—as we describe here—qualitative data are created in real time, avoiding hindsight and recall bias (Golden, 1992).

Verbal protocols are audio-recorded and then transcribed (Ericsson & Simon, 1984, 1993). Since participants report their thoughts without explaining them, it is the researcher’s task to interpret the verbal reports and distill cognitive processes. To guide the researcher’s analysis of verbal reports, qualitative or quantitative data analysis software can be used. Analysis of these reports supports the identification of key variables and, more importantly, the underlying cognitive processes that influence behavior.

BEHAVIORAL EXPERIMENTS FOR CAUSALITY AND REPLICABILITY

The experimental method has helped researchers advance science for centuries. In the 17th century, Isaac Newton used the method to show that white light consists of colors, disproving the common belief that white light was colorless (Newton, 1999). In a series of experiments, he passed white light through a prism, which dispersed the light into a spectrum of colors. To rule out that these colors were created by the prism, he used a second prism and let the red light from the first prism pass through it. Newton observed that the red color remained unchanged, so he concluded that the prism did not create the colors but separated the colors that were already present in the light.

In the 19th century, Louis Pasteur conducted an experiment to demonstrate the effectiveness of his vaccine against anthrax, an infectious disease that was affecting cattle (Stokes, 1997). He compared two groups of animals: one group was vaccinated, and the other served as a control group and was left unvaccinated. One month after the vaccination, the animals were injected with live anthrax bacteria. Only the animals that were vaccinated survived, offering causal evidence that Pasteur’s vaccine protected animals against anthrax.

Such causality is the gold standard in science, and cognition is no exception. In the study of cognition, experiments can assist in establishing causality by allowing researchers to study behavior while controlling for confounding factors (Cook et al., 1979). McGrath (1982) contrasted several methods to conclude that experiments allow researchers to maximize the precision of research findings and draw causal inferences. At the same time, he acknowledges that good researchers weigh external validity to ensure their findings are generalizable. Likewise, Schwenk (1982, p. 214) noted that some (poorly designed) laboratory experiments can seem too artificial, not a good representation of the world outside the laboratory. Over the years, several studies have illustrated how carefully designed experiments that represent decision-making in organizations can provide relevant, generalizable findings (e.g., Ederer & Manso, 2013; Hodgkinson, Bown, Maule, Glaister, & Pearman, 1999; Levine et al., 2017).

In what ways can experiments benefit cognition research? First, researchers can rely on observed, rather than assumed, individual-level inputs and outcomes. In contrast to research that relies on noisy real-life data or proxies, experiments allow for a direct test using clean, observable measures, especially of individual behavior (Croson, Anand, & Agarwal, 2007). Second, individual traits and behavioral outcomes can be measured separately. This removes recall bias (Golden, 1992) and common-method bias (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003), both of which can undermine evidence from interviews or surveys. In contrast, experiments allow for individual traits to be measured separately from decisions, so researchers can assess how one influences the other.

Third, next to fixed individual traits, experiments can capture dynamism, such as evolving experiences. Rather than using static measures, such as those taken at the beginning or a task or after an event happened, in the laboratory participants’ behavior can be measured repeatedly, allowing for insights into longitudinal behavioral patterns, for example, of team decision-making over time (Håkonsson et al., 2016). Finally, experiments allow for easier replication, both in the lab and in the field (Maner, 2016). Once an experimental design has been developed and publicly shared (as it should be), others can use the materials and procedures to collect data in various samples. Such exact replication is more difficult with other methods, such as surveys or case studies, because researchers lack control over various industry- or organization-specific factors that may interfere with research findings (Croson et al., 2007). As mentioned above, replication helps safeguard against false positives and addresses validity concerns, which is why it has become key in many research areas including management (Bettis et al., 2016; Lewin et al., 2016), economics (Camerer et al., 2016), psychology (Open Science Collaboration, 2015), and marketing (Desai, 2013).

COMBINING EXPERIMENTS WITH PROTOCOL ANALYSIS

Each of the methods has strengths and weaknesses—yet in combination they yield unique benefits to students of cognition. Most importantly, they allow researchers to capture cognition alongside realized behavior, in a way that is replicable. Ericsson and Simon (1998, p. 181) advise that for protocol analysis to be successful, participants should be focused on the task at hand without distractions. A controlled lab environment offers such conditions, supporting the collection of reliable protocols.

When combined with protocol analysis, experiments can become a platform for theory building (Colquitt, 2008); they are useful even when sufficient exploratory research is not yet available (Schwenk, 1982). In an exploratory laboratory study, researchers can collect verbal protocols alongside experimental data to gain insights into a range of cognitive processes, thereby supporting theory building. Collecting data from verbal protocols, followed by series of experiments at separate points in time, is an example of combined theory building and theory testing. Specific variables of interest that are identified during protocol analysis can be tested by conducting experiments with or without protocol analysis, at various points in time and in different samples. Combining exploratory and confirmatory methods is a disciplined manner of conducting research (Fiske, 2016) and prevents questionable practices such as hypothesizing after the results are known (HARKing; Kerr, 1998) or selective reporting (John, Loewenstein, & Prelec, 2012; Levine, 2012).

Putting Behavioral Experiments and Protocol Analysis to Work

To illustrate, we describe combining protocol analysis with experiments to concurrently identify cognitive processes and observe their effect on behavior. We elaborate on our research process, from the design of experimental instruments to the replication of research findings.

DESIGNING RIGOROUS EXPERIMENTAL INSTRUMENTS

To benefit from the unique advantages of experiments, the right use of experimental instruments is crucial. To show that white light was not colorless, Newton used prisms as an experimental instrument to observe how light disperses into colors. In cognition research, experimental instruments can come in many forms. As Anderson, Herriot, and Hodgkinson (2001) recommend, both rigor and relevance are key in the design of these instruments. To achieve both, the instrument should carefully match the theoretical features of the studied phenomenon while reflecting managerial decision-making in organizations.

In one study, we were interested in documenting differences in people's behavior under ambiguity and explaining them: Was it luck (Denrell, Fang, & Liu, 2014)? Emotions (Ashkanasy, Humphrey, & Huy, 2017)? Risk aversion (Kahneman & Tversky, 1979)? Such questions would be difficult to answer in organizational settings, where issues such as endogeneity, hindsight bias, and social desirability bias lurk. So, we created an experimental task that simulates managerial decision-making, complete with an ambiguous environment, performance feedback, and iterative decisions. To a participant, the task was a realistic-looking landscape, drawn on a tabletop board, and said to contain hidden oil fields (Fig. 1). The participant's task: Recover as much oil as possible in 20 decisions. Each decision was simply where to drill. After each decision, the participant learned how many oil reservoirs she recovered in a chosen drilling location. Then, the participant chose again and decided between drilling in a known location, drilling in a nearby region, or jumping to an unexplored part of the landscape. From the instructions (Appendix), participants knew that the oil reservoirs tend to be clustered, so the participant was aware that when she chose to stay in a nearby area, she could expect similar amounts of oil.

To design the task, we studied decision-making under uncertainty (Alchian, 1950; Knight, 1921), search on rugged landscapes (Billinger, Stieglitz, & Schumacher, 2013; Kauffman & Levin, 1987), and the exploration–exploitation trade-off (Lavie, Stettner, & Tushman, 2010; March, 1991). Per Knight (1921), the task was ambiguous because a participant had only limited information about the outcomes: She did not know how the oil was distributed or what the maximum or minimum oil availabilities were. Thus, there was no optimal approach (Alchian, 1950)—the participant could only learn from past experiences. The distribution of oil reservoirs mirrored a rugged landscape (Kauffman & Levin, 1987), with peaks of high oil availabilities that were separated by valleys of low availabilities. Throughout the task, the participant faced the exploration–exploitation trade-off, described by March (1991): Sticking with the tried and true by drilling in a known location or pursuing the new and promising, by searching for a new location. As March (1991, p. 71) formulated, the trade-off arises because decision-makers are constrained in time and resources, so they cannot pursue both strategies at once. Similarly, in our task, the participant only had 20 rounds to drill and could not explore and exploit at the same time, but each round chose on a continuum between exploration and exploitation—a design following the suggestions of Lavie et al. (2010) and Mehlhorn et al. (2015).

The experimental design matched theoretical characteristics and has proven its practical relevance: It reflects decision-making in countless managerial situations. For example, our experimental task aptly evokes the tension managers face when they choose between the following range of options: repeating a successful strategy or initiating strategic change; working with the same partners or forming new partnerships; and marketing current products or investing in the development of novel ones. Apart from its managerial relevance, the task also enjoys high face validity (Nevo, 1985): its purpose is clear and easy to understand for participants.

We created a second task to examine how decision-makers' preferences for risk—and separately—ambiguity (also known as uncertainty) affect behavior in the oil drilling task. We captured these preferences in a second instrument so that we could separate measures of individual characteristics, behavior, and performance, avoiding common-method bias (Podsakoff et al., 2003). Starting with a task used in a neuro-economics study (Hsu, Bhatt, Adolphs, Tranel, & Camerer, 2005), we revised it into a web-based instrument (Fig. 2). In it, a participant made 48 decisions; for each, she chose between a sure but low pay, and a higher but risky or ambiguous one. The instrument was designed following Knight's (1921) taxonomy, distinguishing between risk and ambiguity. In risky situations, probabilities of outcomes are known; in ambiguous situations, they are not. So, in the risk treatment, participants knew the probabilities of receiving a higher payment. In the ambiguity treatment, such probabilities were unknown.

Fig. 2 shows a sample screen of the kind of choices participants made. With each choice, a participant could choose a red or blue card — or she could choose to receive a sure payment. Above the cards, a participant saw how many cards were in each deck. Below the cards appeared the number of tokens they would earn if they chose a color that matched the color of a randomly drawn card. On the right, the participant could see how much she would earn if she chose to receive a sure payment.

The example (Fig. 2(a)) shows two decks, one of 10 red cards and one of 10 blue cards. If the color chosen by the participant matched the color of a randomly drawn card, she would earn 10 tokens. If the color the participant chose differed, she would earn nothing. If the participant chose a sure payment, she would earn three tokens. In the ambiguity example (Fig. 2(b)), there are 20 cards, some of them red and some of them blue, but the participant did not know how many of each color. Again, the participant chose between red, blue, or a payment for sure. In the risk example, the participant knew the number of red and blue cards in the deck, so the probabilities of receiving a higher payment were knowable. In the ambiguity treatment, participants only knew the total number of cards in the deck, with no way of knowing the probability of receiving a higher payment.

As part of the design of experimental instruments, researchers should also decide on an appropriate incentive scheme for participants. For their participation, participants may receive monetary incentives, such as cash, or non-monetary incentives, such as a custom report that delineates a participant's characteristics based on her decisions. We chose to offer participants a monetary incentive. We did not offer a fixed fee but used an induced value approach (Smith, 1976), where participants' decisions were paid depending on the decisions they made. In the oil-drilling task, the amount of oil participants accumulated was converted to money. In the risk and ambiguity task, participants were paid for their choice in one randomly drawn round. Because participants' decisions directly affected their ultimate pay-off, participants were more likely to think carefully about the task to maximize their performance.

Selecting an appropriate incentive scheme merits attention, because it may influence participants' behavior in a task. For example, when participants are paid depending on their performance, they may take fewer risks than when they receive a fixed fee, to avoid harming their pay-off (Ederer & Manso, 2013). Researchers therefore should make sure the incentive scheme is aligned with their study purpose.

PREPARING FOR DATA COLLECTION

After designing experimental instruments, researchers can start preparing for data collection. First, experimental materials including the instruments, recruitment messages, and consent form should be submitted for approval to an institutional review board or equivalent ethic committees.

Before data collection starts, it is helpful to develop an experimental protocol: a document that provides a step-by-step overview of the experimental procedure. The protocol ensures that data is collected in a standardized way to facilitate the replication of the experimental procedure at another point in time or at different research sites. The protocol should therefore contain sufficient details so that other researchers can independently execute the same experimental procedure.

COLLECTING DATA IN A CONTROLLED ENVIRONMENT

Once ethical approval is obtained and a detailed experimental protocol is available, researchers can recruit participants and invite them to a laboratory session. Experimental sessions often take place in a computer laboratory, where dozens of participants can simultaneously take part in an experimental task, individually or in groups. When experiments are combined with protocol analysis, sessions should take place in a quiet location, with only one participant per session, to facilitate the "think aloud" process.

Our experimental sessions took place in a conference room, one participant at a time. As the participant arrived, she signed an attendance sheet and read consent information. After obtaining consent, the experiment started. To avoid order effects, each participant completed both experimental tasks in a random order. The participant read instructions and answered comprehension questions. If a participant failed at least one of these questions more than once, she was dismissed without participating. In our experience, 20–35% of the participants fail comprehension questions. Including these questions therefore is crucial to avoid noise in the data.

Before the task started, the participant was asked to "think out loud" during the experiment. As recommended by Ericsson and Simon (1984), the participant first did an exercise to familiarize herself with the process. The experimenter announced: "During this part of the experiment we ask you to think out loud as you are making your decisions. First, let's do an exercise: Please look around the room and say out loud what you are thinking right now." Once the participant felt comfortable verbalizing her thoughts, the experimenter confirmed: "just do the same during the experiment and say out loud whatever you are thinking."

During the experiment, the experimenter reminded the participant: "please think out loud," "feel free to say what is on your mind," and "what are you thinking right now?" At the end of the task, the participant reflected on her overall experience during the task by answering open-ended questions: "What was your overall impression?" and "How did you try to make decisions?" With participants' consent, all sessions were audio-recorded.

When a participant asked questions about the content of the experiment, she was referred to the instructions but not given any additional information. Immediately after the experiment, the participant was paid and confirmed her participation by signing a check-out form. After the session, the experimenter recorded all questions and events, however minute, in an experiment log. This included all questions from participants, any other verbal exchanges with the participants that deviated from the experimental protocol, and all other events, such as when a participant showed up late, left the room during the task, or disagreed regarding her payment.

ANALYZING RICH QUANTITATIVE AND QUALITATIVE DATA

The combination of objective experimental data with subjective thought processes yields a rich data set of qualitative and quantitative data. In our case, the data set included 1,140 drilling decisions and 2,736 risky, ambiguous, or sure choices made by 57 individuals. Transcriptions of participants' verbal reports resulted in 14,563 words, an average of 256 words per participant.

Because we collected data in a controlled environment, we could create clean, observable measures of participants' behavior. For example, our outcome of interest was search behavior. For each drilling spot a participant chose, we calculated the Euclidean distance between that drilling spot and the previous drilling spot as a measure of exploration–exploitation. Low values indicate that a participant chose to exploit the local neighborhood where expected outcomes were similar, whereas high values indicate exploration, a jump to an area where outcomes were unknown. To measure participants' preferences for risk, and separately for ambiguity, we created a profile for each participant based on the repeated decisions she made in the risk and ambiguity instrument.

Next to these objective measures, we analyzed participants' subjective thought processes. To assess the validity of the protocols as data, we checked participants' concurrent and retrospective thought processes for inconsistencies. Because we collected thought processes alongside realized behavior, we could also directly assess inconsistencies between the two. To interpret and explain thought processes, we manually coded each transcript. Following recommended qualitative coding techniques (Miles & Huberman, 1984), we first read through each transcript and created codes that closely matched the words participants used when they reported their thoughts.

Then, we grouped these codes and created overarching codes. For example, codes such as "makes sense," "method," and "pattern," were grouped into the higher-level code "logic." Codes such as "play safe," "bet," and "harm" were grouped under "risk." Qualitative data analysis software like Nvivo is useful to code the data and create visual coding maps (Nvivo, 2012).

Table 1 illustrates the variety of cognitive dimensions that can emerge from the qualitative analysis of verbal protocols. Hodgkinson and Healey (2011) categorize cognition along two dimensions, ranging from intuitive to deliberative and from low to high affect. From the analysis of participants' thought processes, we could capture cognition along these dimensions. For example, we found that participants recognized the role of luck and made intuitive decisions based on lucky numbers or a gist of where they may find oil. Most decisions participants made were more deliberative. Participants often referred to the risk of going to a new region in the landscape, which is why some chose to stay close to already discovered spots. Participants also spoke of rewards, deliberately assessing the prospects of higher rewards or already accumulated rewards to make decisions. Some participants used a form of logic to make decisions. They perceived their decisions as logical and made them in a systematic way, for example, by attempting to establish a pattern of oil availabilities. Although we did not instruct participants to describe their emotions, some nonetheless expressed emotions such as happiness or sadness as they were deciding and receiving feedback.

To further grasp these cognitive dimensions, we used the quantitative text-analysis software Linguistic Inquiry and Word Count (LIWC; Pennebaker, Booth, Boyd, & Francis, 2015). The software analyzes text documents and reports how a document scores on over 90 pre-defined, validated categories such as analytical thinking and emotional tone. The developers first created the software in 1993 and validated and updated it over the years, adding new words and measures (Pennebaker, Boyd, Jordan, & Blackburn, 2015). The software contains an internal dictionary of categories, each consisting of a number of words and word stems that reflect these categories. To calculate output scores, the software analyzes each word in a text document and looks for a match between that word and the words in its internal dictionary. If there is a match, the output score for the category that word represents is incremented. The total output score for a category shows the percentage of words in a text document that matches that category. LIWC also reports summary variables whose scores are calculated in comparison to standardized scores from the developers' earlier published work (e.g., Cohn, Mehl, & Pennebaker, 2004; Pennebaker, Chung, Frazee, Lavergne, & Beaver, 2014).

The software is useful to capture the deliberate side of cognition, but it also contains output scores that capture affect, such as the summary variable "emotional tone," which reflects the positive emotion words (e.g., "good," "nice") and negative emotion words (e.g., "bad," "angry"). As for the deliberate cognitive dimensions, the software reports output scores for the categories of risk and rewards. The scores indicate to what extent participants use words that match the risk and reward categories in the software. The risk category contains 103 words such as "danger" and "doubt." The rewards category contains 120 words, including "take," "prize," and "benefit." To capture to what extent participants relied on logic when making decisions, the software reports a summary variable "analytical thinking." The LIWC scores can be used to complement or validate manual coding, supporting the identification of various dimensions of cognition.

REPLICATING RESEARCH FINDINGS IN VARIOUS SAMPLES

By adding cognitive dimensions derived from protocol and text analysis to the observation of realized behavior in an experiment, researchers can grasp cognition in action. Then the same experimental procedure can be replicated to test the validity of research findings.

As we illustrated, the physical board that we used during the oil-drilling task helped us identify and examine important decision-making processes underlying behavior in an ambiguous setting. The instrument is both inexpensive to create and easy to set up, allowing for easy replication in various settings, both in the lab and in the field. These features are particularly beneficial to replicate in populations that are not western, educated, industrialized, rich, or democratic (WEIRD), populations about which we know far too little (Henrich, Heine, & Norenzayan, 2010).

Collecting experimental data using protocol analysis, with only one participant per session, can be time consuming. So, to facilitate replication—and extension—of these findings on a larger scale, we created a web-based version of the task, which was used without protocol analysis. This way, the mechanisms identified during lab sessions with protocol analysis can be tested in larger samples, as in non-student samples in artefactual field experiments (Harrison & List, 2004; Levitt & List, 2009) or online labor markets, such as Amazon Mechanical Turk, CrowdFlower, and Prolific Academic (Horton, Rand, & Zeckhauser, 2011; Peer, Brandimarte, Samat, & Acquisti, 2017).

Conclusion

We proposed that combining experiments with protocol analysis offers benefits that can advance the study of managerial and organizational cognition. We illustrated how this combination of methods can help researchers identify and examine a range of cognitive processes that can influence behavior in ambiguous settings. The findings can then be replicated and extended across populations, easily and cheaply. We suggest that this mix of methods offers a solution to three methodological challenges: capturing underlying processes, establishing causality, and ensuring validity. While experiments allow for the study of realized behavior and its causal effects, protocol analysis is a useful technique to simultaneously probe into underlying decision-making processes. This set of benefits is difficult to achieve using other methods. We suggest that experiments and protocol analysis can advance cognition research in important ways, especially to examine cognition in action, in a way that is replicable.

Note

1. Protocol analysis is related to "Think Aloud," which Laureiro-
Martinez discusses in another chapter in this volume. Here we adopt the term used by Ericsson and Simon (1984).

Acknowledgments

We thank audiences at the Organization Science Winter Conference, at the MOC-TIM meeting at ETH Zürich, and at meetings of the Academy of Management, the Strategic Management Society, the Association for Psychological Science, and the Economic Science Association. SSL and CR acknowledge a grant from the European Research Council (695256). SSL also received a grant from the Hong Kong Research Council General Research Fund (14655416).

References

Agarwal, R., Croson, R., & Mahoney, J. T. (2010). The role of incentives and communication in strategic alliances: An experimental investigation. *Strategic Management Journal*, *31*(4), 413–437. doi:10.1002/smj.818

Alchian, A. A. (1950). Uncertainty, evolution, and economic theory. *The Journal of Political Economy*, *58*(3), 211–221.

American Psychological Association (2009). Cognition. In *Glossary of Psychological Terms*, Retrieved March 5, 2017 from http://www.apa.org/research/action/glossary.aspx?tab=3

Anderson, N., Herriot, P., & Hodgkinson, G. P. (2001). The practitioner–researcher divide in industrial, work and organizational (IWO) psychology: Where are we now, and where do we go from here? *Journal of Occupational and Organizational Psychology*, *74*(4), 391–411.

Ashkanasy, N., Humphrey, R., & Huy, Q. (2017). Integrating emotions and affect in theories of management. *Academy of Management Review*, *42*(2), 175–189.

Baron, R. A., & Ensley, M. D. (2006). Opportunity recognition as the detection of meaningful patterns: Evidence from comparisons of novice and experienced entrepreneurs. *Management Science*, *52*(9), 1331–1344.

Begley, C. G., & Ellis, L. M. (2012). Drug development: Raise standards for preclinical cancer research. *Nature*, *483*(7391), 531–533.

Bergh, D. D., Sharp, B. M., Aguinis, H., & Li, M. (2017). Is there a credibility crisis in strategic management research? Evidence on the reproducibility of study findings. *Strategic Organization*, *15*(3), 423–436.

Bettis, R. A., Helfat, C. E., & Shaver, J. M. (2016). The necessity, logic, and forms of replication. *Strategic Management Journal*, *37*(11), 2193–2203. doi:10.1002/smj.2580

Billinger, S., Stieglitz, N., & Schumacher, T. R. (2013). Search on rugged landscapes: An experimental study. *Organization Science*, *25*(1), 93–108.

Bruce, N., & Kristine, B. (2008). Good to Great, or just good? *The Academy of Management Perspectives*, *22*(4), 13–20.

Butler, N., Delaney, H., & Spoelstra, S. (2017). The gray zone: Questionable research practices in the business school. *Academy of Management Learning & Education*, *16*(1), 94–109. doi:10.5465/amle.2015.0201

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., …, Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, *351*(6280), 1433–1436. doi:10.1126/science.aaf0918

Chang, A. C., & Li, P. (2015). *Is economics research replicable? Sixty published papers from thirteen journals say "usually not"*. Washington, DC: Board of
Governors of the Federal Reserve System.

Clark, D. R., Li, D., & Shepherd, D. A. Country familiarity in the initial stage of foreign market selection. *Journal of International Business Studies*. Advance Online Publication.

Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2004). Linguistic markers of psychological change surrounding September 11, 2001. *Psychological Science*, *15*(10), 687–693.

Colquitt, J. A. (2008). From the editors publishing laboratory research in AMJ: A question of when, not if. *Academy of Management Journal*, *51*(4), 616–620. doi:10.5465/amr.2008.33664717

Cook, T. D., Campbell, D. T., & Day, A. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Boston, MA: Houghton Mifflin.

Croson, R., Anand, J., & Agarwal, R. (2007). Using experiments in corporate strategy research. *European Management Review*, *4*(3), 173–181.

Denrell, J., Fang, C., & Liu, C. (2014). Perspective—Chance explanations in the management sciences. *Organization Science*, *26*(3), 923–940.

Desai, P. (2013). Marketing science replication and disclosure policy. *Marketing Science*, *32*(1), 1–3.

Dew, N., Read, S., Sarasvathy, S. D., & Wiltbank, R. (2009). Effectual versus predictive logics in entrepreneurial decision-making: Differences between experts and novices. *Journal of Business Venturing*, *24*(4), 287–309.

Ederer, F., & Manso, G. (2013). Is pay for performance detrimental to innovation? *Management Science*, *59*(7), 1496–1513.

Eich, E. (2014). Business not as usual. *Psychological Science*, *25*(1), 3–6.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: The MIT Press.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Revised ed.). Cambridge, MA: MIT Press.

Ericsson, K. A., & Simon, H. A. (1998). How to study thinking in everyday life: Contrasting think-aloud protocols with descriptions and explanations of thinking. *Mind, Culture, and Activity*, *5*(3), 178–186.

Fiske, S. T. (2016). How to publish rigorous experiments in the 21st century. *Journal of Experimental Social Psychology*, *66*, 145–147.

Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin*, *137*(2), 316.

Golden, B. R. (1992). Research notes. The past is the past – Or is it? The use of retrospective accounts as indicators of past strategy. *Academy of Management Journal*, *35*(4), 848–860.

Goldfarb, B., & King, A. A. (2016). Scientific apophenia in strategic management research: Significance tests & mistaken inference. *Strategic Management Journal*, *37*(1), 167–176.

Gould, S. J. (1999). *Protocol and cognitive response analysis*. Northampton, MA: Cheltenham.

Grégoire, D. A., Barr, P. S., & Shepherd, D. A. (2010). Cognitive processes of opportunity recognition: The role of structural alignment. *Organization Science*, *21*(2), 413–431.

Håkonsson, D. D., Eskildsen, J. K., Argote, L., Mønster, D., Burton, R. M., & Obel, B. (2016). Exploration versus exploitation: Emotions and performance as antecedents and consequences of team decisions. *Strategic Management Journal*, *37*, 985–1001.

Hanfmann, E., & Kasanin, J. (1937). A method for the study of concept formation. *The Journal of Psychology*, *3*(2), 521–540.

Harrison, G. W., & List, J. A. (2004). Field Experiments. *Journal of Economic Literature*, *42*(4), 1009–1055. doi:10.1257/0022051043004577

Helfat, C. E., & Peteraf, M. A. (2015). Managerial cognitive capabilities and the microfoundations of dynamic capabilities. *Strategic Management Journal*, *36*(6), 831–850. doi:10.1002/smj.2247

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, *466*(7302), 29.

Hodgkinson, G. P., Bown, N. J., Maule, A. J., Glaister, K. W., & Pearman, A. D. (1999). Breaking the frame: An analysis of strategic cognition and decision making under uncertainty. *Strategic Management Journal*, *20*, 977–985.

Hodgkinson, G. P., & Healey, M. P. (2008). Cognition in organizations. *Annual Review of Psychology*, *59*(1), 387–417. doi:10.1146/annurev.psych.59.103006.093612

Hodgkinson, G. P., & Healey, M. P. (2011). Psychological foundations of dynamic capabilities: Reflexion and reflection in strategic management. *Strategic Management Journal*, *32*(13), 1500–1516.

Hodgkinson, G. P., & Sparrow, P. R. (2002). *The competent organization: A psychological analysis of the strategic management process*. Buckingham: Open University Press.

Horton, J., Rand, D., & Zeckhauser, R. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics*, 1–27. doi:10.1007/s10683-011-9273-9

Hsu, M., Bhatt, M., Adolphs, R., Tranel, D., & Camerer, C. F. (2005). Neural systems responding to degrees of uncertainty in human decision-making. *Science*, *310*(5754), 1680–1683.

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Med*, *2*(8), 696–701.

Isenberg, D. J. (1986). Thinking and managing: A verbal protocol analysis of managerial problem solving. *Academy of Management Journal*, *29*(4), 775–788. doi:10.2307/255944

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*(5), 524–532.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, *47*(2), 263–292.

Kauffman, S., & Levin, S. (1987). Towards a general theory of adaptive walks on rugged landscapes. *Journal of Theoretical Biology*, *128*(1), 11–45.

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, *2*(3), 196–217. doi:10.1207/s15327957pspr0203_4

Knight, F. H. (1921). *Risk, uncertainty and profit*. New York, NY: Hart, Schaffner and Marx.

Lavie, D., Stettner, U., & Tushman, M. L. (2010). Exploration and exploitation within and across organizations. *Academy of Management Annals*, *4*(1), 109–155. doi:10.1080/19416521003691287

Lee, E., L., & Puranam, P. (2015). The nature of expertise in organization design: Evidence from an expert–novice comparison. In G. Gavetti & W. Ocasio (Eds.), *Cognition and Strategy (Advances in Strategic Management Volume 32)*
(pp. 181–209). Emerald Group Publishing Limited.

Levine, S. S. (2012). Walter R. Nord and Ann F. Connell: Rethinking the knowledge controversy in organization studies: A generative uncertainty perspective. *Administrative Science Quarterly*, *57*(3), 537–540. doi:10.1177/0001839212462542

Levine, S. S., Bernard, M., & Nagel, R. (2017). Strategic intelligence: The cognitive capability to anticipate competitor behavior. *Strategic Management Journal*, Advance online publication.

Levitt, S. D., & List, J. A. (2009). Field experiments in economics: The past, the present, and the future. *European Economic Review*, *53*(1), 1–18. doi:10.1016/
j.euroecorev.2008.12.001

Lewin, A. Y., Chiu, C.-Y., Fey, C. F., Levine, S. S., McDermott, G., Murmann, J. P., & Tsang, E. (2016). The critique of empirical social science: New policies at management and organization review. *Management and Organization Review*, *12*(4), 649–658.

Li, H., Daugherty, T., & Biocca, F. (2001). Characteristics of virtual experience in electronic commerce: A protocol analysis. *Journal of Interactive Marketing*, *15*(3), 13–30.

Maner, J. K. (2016). Into the wild: Field research can increase both replicability and real-world impact. *Journal of Experimental Social Psy-*

chology, *66*, 100–106.

March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, *2*(1), 71–87.

McGrath, J. (1982). Dilemmatics: The study of research choices and dilemmas. In J. McGrath, J. Martin, & R. Kulka (Eds.), *Judgment Calls in Research*
(pp. 69–102). Beverly Hills, CA: Sage.

Mehlhorn, K., Newell, B. R., Todd, P. M., Lee, M. D., Morgan, K., Braithwaite, V. A., …., Gonzalez, C. (2015). Unpacking the exploration–exploitation tradeoff: A synthesis of human and animal literatures. *Decision*, *2*(3), 191–215. doi:10.1037/dec0000033

Miles, M., & Huberman, A. (1984). *Qualitative data analysis: A sourcebook of new methods*. Beverly Hills, CA: Sage Publications.

Mill, J. S. (1883). *A system of logic ratiocinative and inductive: Being a connected view of the principles of evidence and the methods of scientific investigation*. London: Longmans.

Nadkarni, S., & Chen, J. (2014). Bridging yesterday, today, and tomorrow: CEO temporal focus, environmental dynamism, and rate of new product introduction. *Academy of Management Journal*, *57*(6), 1810–1833.

Nevo, B. (1985). Face validity revisited. *Journal of Educational Measurement*, *22*(4), 287–293.

Newton, I. (1999). *Philosophiae naturalis principia mathematica, general scholium* (I. B. Cohen & A. Whitman, Trans. 3rd ed.). Berkeley and Los Angeles, CA: University of California Press.

QSR International. (2012). NVivo qualitative data analysis software 10.

Nvivo qualitative data analysis Software; QSR International Pty Ltd. Version 10, 2012

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251).

Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior and Human Performance*, *16*(2), 366–387.

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, *70*, 153–163. doi:https://doi.org/10.1016/j.jesp.2017.01.006

Pennebaker, J. W., Booth, R. J., Boyd, R. L., & Francis, M. E. (2015). *Linguistic inquiry and word count: LIWC2015*. Austin, TX: Pennebaker Conglomerates.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. Austin, TX: University of Texas at Austin.

Pennebaker, J. W., Chung, C. K., Frazee, J., Lavergne, G. M., & Beaver, D. I. (2014). When small words foretell academic success: The case of college admissions essays. *PloS ONE*, *9*(12), e115844.

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.

Popper, K. R. (1959). *The logic of scientific discovery*. New York, NY: Basic Books.

Popper, K. R. (1963). *Conjectures and refutations; the growth of scientific knowledge*. London: Routledge and Kegan Paul.

Robie, C., Brown, D. J., & Beaty, J. C. (2007). Do people fake on personality inventories? A verbal protocol analysis. *Journal of Business and Psychology*, *21*(4), 489–509.

Schwenk, C. R. (1982). Why sacrifice rigour for relevance? A proposal for combining laboratory and field research in strategic management. *Strategic Management Journal*, *3*(3), 213–225.

Simmel, M. L. (1953). The coin problem: A study in thinking. *The American Journal of Psychology*, *66*(2), 229–241.

Smith, V. L. (1976). Experimental economics: Induced value theory. *American Economic Review*, *66*(2), 274–279.

Sommer, S. C., & Loch, C. H. (2004). Selectionism and learning in projects with complexity and unforeseeable uncertainty. *Management Science*, *50*(10), 1334–1347.

Stokes, D. E. (1997). *Pasteur's quadrant – Basic science and technological innovation*. Washington, DC: Brookings Institution Press.

Tsang, E. W. K., & Yamanoi, J. (2016). International expansion through start-up or acquisition: A replication. *Strategic Management Journal*, *37*(11), 2291–2306. doi:10.1002/smj.2569

Appendix: Drill for Oil! Earn Money!

In front of you is a large lot of land, represented by a board that is covered by identically colored (blue) notes. You are interested in what lies beneath—oil reservoirs—which are like underground lakes of oil, stretching over an area in a pattern. You will choose drilling locations to reach those reservoirs and capture as many oil barrels as possible.

Some of the oil reservoirs may lay closer to the surface, easier to reach, and therefore more profitable. Some of them may lay deep underground, requiring more effort, and therefore less profitable. Some areas may be completely dry.

Just by looking at the sand, you cannot know where oil reservoirs lie. The only way to find oil is by drilling. You will have the opportunity to drill once per round for 20 rounds. The more oil you find, the more money you will earn.

In each round, choose a spot for drilling and place the marker on the spot. For example, if you want to drill in spot E19, place the marker on the intersection of row E and column 19. After you decide where to drill and place the marker there, the experimenter will reveal the outcome. The color under the sand indicates the profitability of drilling in that spot. We will give you a card showing the number of barrels you earned from the chosen spot.

Remember that the reservoirs may form a pattern, so the amount of oil in one spot may be related to the amount of oil in its neighboring spots. For example, imagine that you first drilled in the bottom-left corner of the board. The spot above it may have a similar amount of oil, but not always.

In each round, you may choose a different spot. You may also keep drilling in the same spot, which means that you will earn the same number of barrels.

The locations and sizes of oil reservoirs are set before the game begins. During the game, the locations and the sizes are not affected by anyone's actions and do not change.

At the end of the game, the number of oil barrels you found will be converted to cash at a predetermined rate and paid to you.

Fig. 1: An Oil Drilling Task.

Fig. 2: A Measure of Risk (a) and Ambiguity (b).

Table 1: Illustrative Cognitive Dimensions Derived from the Qualitative Analysis of Participants' Verbal Reports.

Cognitive Dimensions	
Luck	I will see my luck for a couple of rounds and see if I can get some more oil. How about drilling location D13, because that is my lucky number and my name starts with a D.
Risk	Considering that there is no oil on the top, I don't want to take much of a risk going there, a safer position would be J11. Finding a new oil location would be a risk, and I'd ra-ther reap the benefit of what I've already found, so I'm thinking of digging again in the same spot.
Rewards	Now I want to take a chance because I've earned a lot of money so I want to go a bit further, but not much, J11. We can keep drilling in the same spot which means that I can earn the same amount of oil, so I would like to go with H11. So that my points increase, I would again go with I11.
Logic	This area seems to have a higher concentration of oil; I have three red spots, so the most logical choice should be O14. It is like a minesweeper game right now, oh nice, I got a good spot, so now F10, because logically it makes senses, to make sure that I am covering up the whole area, so that is done.
Emotions	I'm thinking that I'm having fun, is that ok? I am trying to understand how this works. I'm thinking that I'm happy I found a place that is more than 400.