Find Out A New Method to Study Abductive Reasoning in Empirical Research

Dorota Żelechowska¹, Natalia Żyluk¹, and Mariusz Urbański¹

Abstract
This article presents a new tool that provides a methodological context to observe and analyze, both qualitatively and quantitatively, manifestations of abductive reasoning in empirical research. Abduction is a form of a complex reasoning carried out to make sense of surprising or ambiguous phenomena or fill the gaps in our beliefs. Despite the ubiquity of abduction in professional and everyday problem-solving processes, little empirical research was dedicated to investigate this type of reasoning, and most of them focused on products of abduction—abductive hypotheses. Our instrument, Find Out, catches abduction as a real-life form of reasoning consisting of two phases—generation and evaluation of hypotheses. It offers the possibility to account on abduction from both product and process perspective and enables both qualitative and quantitative analyses on gathered data to be conducted. In this article, the task and examples of qualitative analyses of the data are presented.

Keywords
abductive reasoning, hypotheses generating, hypotheses testing, explaining, making sense

Aim of the Study
In this article, we describe a new methodological tool, called Find Out, which enables a researcher to observe and capture the process of abductive reasoning. To justify the decisions concerning the form and content of the task, not only the final version of the method but the process of constructing and validating it are briefly presented. The task collects data relating to both the products and the process of abductive reasoning involving two phases—hypotheses generation and hypotheses testing. An effort was put into designing a context that would allow the gathering of data to be suitable for not only quantitative but also qualitative analyses. Some examples of analyses are presented in the final part of the article. Our aim was to design a paradigm that, however complex, would make it possible to identify more variables indicating individual differences in performance than simple measures related to the correctness of explanations. As we could not find a satisfying task, we had to design it by ourselves.

Introduction
Abductive Reasoning
Abductive reasoning serves the purpose of making sense of surprising, ambiguous, or otherwise puzzling phenomena in order to fill the gaps in our beliefs, maintaining or restoring their coherence (Thagard, 2000; Thagard & Shelley, 1997). This type of reasoning is ubiquitous in both professional contexts such as scientific research, forensic investigation, clinical diagnose, and everyday ones such as natural language understanding, empathy, and theory of mind (Hobbs, 2006; Otten et al., 1995; see, e.g., Magnani, 2009). Surprisingly enough, it is also quite mysterious. There is widespread agreement that the essence of abduction is captured by the schema proposed by Peirce (CP, 5.189):

The surprising fact, C, is observed.
But if A were true, C would be a matter of course.
Hence, there is reason to suspect that A is true.

However, a number of questions arise concerning the details of this schema (see Woods, 2017). Our theoretical framework was set up in line with an instrumental account on abduction (Urbański & Klawiter, 2018). We interpret abduction as a

¹ Faculty of Psychology and Cognitive Science, Adam Mickiewicz University, Poznań, Poland

Corresponding Author:
Dorota Żelechowska, Faculty of Psychology and Cognitive Science, Adam Mickiewicz University, Szamarzewskiego 89/AB, Poznań 60-568, Poland.
Email: dorota.zelechowska@gmail.com
complex form of reasoning (Ajdukiewicz, 1974), comprising both generation and evaluation of hypotheses (Schurz, 2017) aimed at selection of the best or most promising of them (which does not imply that there is only one such hypothesis). An exemplification of such an account is the model of an application of abductive reasoning to clinical practice in psychology by Vertue and Haig (2008). Their model assumes the following steps of abductive reasoning: (a) phenomena detection (first, data collection then data analysis), (b) inferring causal mechanisms, (c) developing a causal model, (d) evaluating the causal model, and (e) formulating the case. In this account, generation and evaluation of hypotheses are intertwined and not easily separable as even the analysis of data requires some prior assessment of what is prospective and what is not. This is very close to our interpretation of abduction. However, as our aims are explorative rather than confirmative in regard to establishing a model of abductive reasoning, we are not going to hypothesize on what are the consecutive steps in abductive reasoning at this stage of our research.

Another source worth referring to is also Tavory and Timmermans (2014). The authors’ interest in abduction is on a methodological level rather than an object level as they are concerned with abduction as a qualitative data analysis approach, not accounting for abductive reasoning per se. Nevertheless, their understanding of what abduction consists in is very much in line with ours—a concise witness to that is their “Synopsis of abductive analysis” (Tavory & Timmermans, 2014, pp. 131–132).

**Known Measures of Abductive Reasoning**

According to our experience, finding a task capable of evaluating abductive reasoning in empirical research is a challenge. An example of a method that enables a researcher to analyze the process of abductive reasoning in both quantitative and qualitative approaches was used by Oh (2008). In his research, empirical data referring to the process of learning were analyzed based on the framework provided by the abductive inquiry model. According to that model, abductive reasoning consists of four phases:

- Exploration
- Examination
- Selection
- Explanation

Oh investigated abductive inquiry as a method that could be used as a helping aid for teachers of earth science. Abductive method appeared to be an effective way to teach scientific reasoning in case of both older and younger students and promoted understanding of complicated phenomena. After analyzing the process of learning by abductive inquiry, Oh described what exactly had happened during each phase of the model. During the exploration phase, participants investigated earth scientific phenomena with provided data and transformed them into problems that could be explained abductively. The examination phase needed activation and an expansion of background knowledge to find rules for abductive inference and then some hypotheses were generated (thus, these two phases cover three first parts of the Vertue & Hague, 2008, model). Selection consisted of evaluating hypotheses empirically and/or theoretically and then choosing the most plausible ones. The final, exploration phase resulted in causal and narrative explanations using previously selected hypotheses.

A study referring to those two phases was conducted by Donnelly et al. (1990). They used a measure designed to assess clinical problem-solving. It consisted of three tasks requiring the listing of possible diagnoses that should be tested, starting from the most to the least promising. The second part of the task required a description of the way of testing generated hypotheses. In this study, responses were judged by experts only in terms of their accuracy. Despite that limitation, it seems that further qualitative analyses of data could also be possible.

However, most measures of abductive reasoning seem to limit possible analyses of data to the quantitative approach and relate only to a single phase of the process. For instance, a task used by Mirza (2015) to assess the effects of abductive reasoning training referred only to the ability of generating hypotheses. A method was profiled to the expertise area of the participants (baccalaureate nurse students) and was based on “care scenarios” that presented a situation that required an explanation. The evaluation system was modeled on criteria proposed by Ingram and colleagues (1998): accuracy (the number of correct hypotheses based on the expert list of responses), expertise (the number of not only correct but ideal responses based on the expert list of responses), and breadth (the number of biopsychological areas covered by the responses).

Similarly, a task requiring a hypothesis to be generated was used by Kwon, Lee, Shin, and Jeong (2009) in the research aimed at investigating the learning-related changes in brain activation induced by the training of hypotheses generation skills. The authors assumed that the process of formulating hypotheses consists of six steps:

- Observing a situation
- Generating a causal question
- Analyzing the question
- Representing experienced phenomena
- Causing representation
- Constructing hypotheses

A training task was designed for academic students and required the generation of biological hypotheses (exemplary task question: “Why is the monkey covered with white fur?”). The evaluation system was limited to the final step of formulating hypotheses, that is, they were evaluated by counting the number of explanations included in each of them. This model differs substantially to the ones proposed by Vertue & Hague (2008) and Oh (2008).

A slightly different version of a method to evaluate the hypothesis generation ability, adjusted to abilities of primary school students, was used in a project carried out by Kwon et al.
of two intertwined components, generation and evaluation of research, we addressed this gap. cause by individual differences between participants. In our methodological loophole in the case of the process of abduction (understood as taking into account different aspects of the situation). The findings of the review indicated a major meth-

sional or abstract context. Variables assessed by the measures of abductive reasoning by authors at all. The presented tasks were not usually associated with abductive reasoning or just to the hypothesis generating or testing phase. As consequence, they are not usually associated with abductive reasoning by authors at all. The presented tasks require a rather specific knowledge and are set in a professional or abstract context. Variables assessed by the measures are usually limited to the correctness of the hypotheses (depending on the opinion of experts) and the broadness (understood as taking into account different aspects of the situation). The findings of the review indicated a major methodological loophole in the case of the process of abductive reasoning, especially in terms of the possible variability caused by individual differences between participants. In our research, we addressed this gap.

As we interpret abduction as a reasoning process consisting of two intertwined components, generation and evaluation of hypotheses, a complex task allowing both of them to be addressed was the one we needed. One important constraint was that the context of a task should be close to common, real-life problems. However, in spite of informal content and the structure of the task, the method should maintain requirements of standardization so as to enable further compara-

Method: Find Out Task
To construct the measure of abductive reasoning that would meet our requirements, the following procedure was applied: choosing the most suitable model of abductive reasoning, defining constructional requirements of the task (conceptual and methodological), researching the relevant literature in search of existing methods, creating primary sketches of the task (more than 40 sketches were taken into consideration), developing a chosen project, a pilot study to test the preliminary project of the task, a constructional pilot study to develop the final project of the task, and an exploratory pilot study to analyze the data that the task enables a researcher to gather.

During the process of constructing the task, 12 versions of the method were created—most digital, a few in a paper-pencil format and one that can be conducted as spoken communication. Most of them were updated, and corrected versions of the previous ones and presenting all of them would exceed the limit allowed for this article. In case of any interest in alternative tools, we encourage the reader to contact the authors of the paper directly.

To facilitate the understanding of the task, we start by presenting the final and chosen version of the method. Only then major modifications that have been introduced to the tool are described. Find Out is a computerized task composed of three consequent parts integrated by the background story and the general aim of the task, which is to explain an ambiguous situation. Each part of the task represents a different stage of a process of solving an abductive problem. The subject is asked to play the role of a detective working for a local media outlet.
and who is seeking for an explanation of what has happened in the presented enigmatic situation. She or he has to prepare some information for an upcoming “breaking news” edition. The whole task is planned to take 65 min (with instructions), but the participant can complete the task before the established time limit. The procedure is described below. For the sake of clarity regarding the description of this rather complex task, each part of the task is presented separately, just after the information concerning the general procedure.

**General Procedure**

The test should be conducted individually in laboratory settings with the use of computer equipment. The task is performed with the use of Google Docs—a text document and a chat window. The participant is seated in a separated room in front of a computer. The person conducting the experiment is in a room next door so as to provide the participant with more comfortable conditions of work but, at the same time, be able to help in case of any questions regarding the procedure or any technical difficulties. First, the subject is acquainted with a tool and encouraged to ask any questions. Then, the person conducting the experiment leaves the room and communication starts to be mediated via the chat window. Schematic representation of the task is presented below (Figure 1).

The first message (appearing in the chat window) is sent to encourage the participant to start the task and show the proper way of communicating (online). The procedure consists of three stages, each with a time limit, which are presented one by one. The participant starts the next stage only after completing the former part of the task. Instructions are delivered gradually, for each stage separately so as to encourage the participant to perform only the task required at a certain stage. Each piece of instructional information or reply sent to the participant is highly standardized. The person conducting the experiment is equipped with a list of specially prepared information and copy it over to the major body of the task in a timely manner or when it is necessary. There is also a specially prepared FAQ document with issues and solutions that appeared in the pilot studies and may be problematic. The conductors of the experiment must complete training under the supervision of more experienced researchers before they can start conducting the test.

Research is carried out in Polish, but for the needs of the article, we have translated the instructions and sample performances of one participant to enable readers to have a better understanding of the task. Google Docs with translated instructions and sample performances for each stage of the task can be found online Supplementary Materials. A scheme of the procedure of the investigation is presented in Table 1.

**Part 1: Warm-up.** The first stage of the task requires formulating hypotheses. It takes 20 min and starts with sending a message encouraging the subject to start the task and suggests the proper way of communicating (only via chat). In the Online Supplementary Materials, you can see this particular stage (Find Out Stage 1).

At this moment of the procedure, the story is presented. The content of the story is highly inspired by a story from the Black Story game (by Moses, Polish version by Rebel), and it reads as follows:

On a winter’s night police stopped a car for a routine control. The driver turned out to be a famous celebrity (his identity is now known to the journalists). He had two bottles of mulled wine with him. He told the police that he had bought the wine 40 minutes before from the Christmas market held in the city centre—more specifically, from the stand of the manufacturer of regional specialties, including good-quality mulled wine. The man claimed that he had been going to visit his friends and he had bought the wine as a gift for them. However, it turned out that one of bottles was half empty and a distinct odor of mulled wine was present in the car. The driver admitted that he had succumbed to temptation and had drunk half of the bottle of the wine by himself. The police officers gave him a breathalyzer test and read its result. Subsequently, they had a conversation with the man. Immediately after, the driver drove off. The police officers consulted each other, got into their car and took off in the direction of the city centre.

The subject is asked to write down all ideas that could possibly explain what has happened in a presented situation in a notepad. It requires taking into account everything that is considered by the participant as unclear or unknown. At the top of the screen, there is also a scratch pad where the subject can make any additional notes.

Some major modifications were introduced to the original story in order to provide many possible variants of a viable explanation at different levels of solving the problem. After 20 min, the person conducting the experiment reminds the participant that the time is up and she or he can only finish the sentence. The next stage then begins. An example of one participant’s performance (Subject 1) after the first stage can be found in Online Supplementary Materials (Example: Find Out Stage 1).

**Part 2: Cooperation.** The second stage lasts 30 min. It requires not only generating but also testing previously generated hypotheses and new ideas. Now, the subject works on already devised explanations but is allowed to ask yes–no questions to collaborators via chat. The participant is asked to imagine that the collaborators are competent and can give every necessary piece of information. However, there is one condition that has to be satisfied to receive any answer—the question must be followed by justification pointing out why this particular question is asked. The background story presents the task as cooperation with journalists who are preparing the news story for the upcoming edition. As a consequence, the participant is obliged to update explanations and other notes written in the notepad in accordance with upcoming new information (delete outdated, modify existing, and add new) so as to expedite the work of the journalists. After 30 min, the subject is asked to stop performing this part of the task.
A blank template for the second stage of the task (*Find Out Stage 2*) is presented in the Online Supplementary Materials. The performance of the Subject 1 can also be found in the Online Supplementary Materials. Here you can see modifications introduced to the notepad at the very end of the stage (Example: *Find Out Stage 2*). Recorded content of the chat, including the second stage and cooperation part (that lasted from 14:16 to 14:41), can be found in the Online Supplementary Materials (Example: *Find Out Chat window*).
The final version of the task is based on some basic assumptions (presented above) on the proper methodological context that enables the observation of expressions of abductive reasoning in a context close to real-life conditions and the results of two pilot studies. In order to make it clear why some methodological, procedural, or structural decisions on the task were accepted, the most significant findings of two pilot studies are briefly described below.

Pilot Study Number 1: Testing Preliminary Version of the Task

The major goal of the first study was to investigate whether a task built on a mystery story can act as an adequate methodological context to investigate abductive reasoning in terms of the hypotheses generation phase. It was expected that this measure would enable individual differences among participants to be observed. If so, the next step would be to develop the procedure in the direction that enables not only a hypothesis generation phase to be evaluated but also a testing phase. Forty-five undergraduate students (33 women, 11 men, 1 preferred not to answer) aged 18–22 ($M = 19.58, SD = 0.78$) participated in the study. We used a specially designed paper-pencil test called the hypotheses generation task. The test consists of two similar tasks that require generating possible explanations for a surprising situation presented in the story. Both stories were modified versions of the stories from a game Black Stories (by Moses).

Each story includes several aspects that are underspecified and need to be accounted for. For example, in Story Number 1, it is not known what the result of the breathalyzer test was, what the content of a conversation between the driver and police officers was, and what had really happened that the bottle was half empty. In order to direct the process of formulating explanations, each story ended with a question:

What has happened and did anyone break the law? Both stories used in the first pilot study are presented in Online Supplementary Materials (Find Out Pilot Study 1: Stories).

At that point of our research, we thought that an additional request (i.e., “whether anyone broke the law”) would not only help to better understand the requirements of the task but also to facilitate generating ideas involving different aspects of the situation. Participants were tested in two groups, but each person worked individually. There was no time limit as we wanted to assess how long it would take to complete the test. The instructions included the suggestion that the performance of the task should take no more than 30 min, but participants were informed that anyone who would prefer to work longer would be given the opportunity to do so.

Ten participants shared their opinions on the task. It could be concluded that the participants enjoyed solving the task and that the instructions were clear enough (no comment suggested otherwise). However, the idea to ask about breaking the law did not prove itself as facilitation. Instead, some participants reported feeling incompetent to judge any legal issues. As a result, we have decided to delete the additional question from further versions of the task. It appeared that the first story (about a driver and the police) enabled us to gather observations that demonstrated more individual differences than the second story (about coworkers), which was probably too sketchy and therefore inspired the subjects to create similarly sketchy explanations. The next versions of the task were based on the modification of the first but not the second story.

Further modifications introduced to the task were chosen to accomplish two partially disparate goals: to optimize the chance of observing individual differences and to standardize the context of their appearance to reduce the risk of confounding factors, which could impair the objectivity of future studies on the individual differences in abductive reasoning.

Pilot Study Number 2: Constructing the Final Version of the Task

After analyzing and reconsidering the results of the first pilot study, some decisions were made on the preliminary scheme of the task. The next step was a constructive pilot study conducted to accomplish two major goals: to define the final form of the experimental task and to decide on the details of the procedure of the testing routine. At this stage, the task was being modified just after the people conducting the experiment had received sufficient feedback to improve the method. As a consequence, 12 versions of the task were empirically tested and evaluated. Nineteen subjects (aged 24–69) participated in the study. The group of participants was varied enough to show a wide diversity of performances and enabled us to receive diverse feedback. Each version of the task was performed by one, two, or three participants.

The following major modifications were introduced during this stage: changing the format of the medium from an analog
paper-pencil task to a computerized task. Although the analog version of the task is still available on request and can be useful in some cases (please contact authors in case of interest), the digital version enabled us to shorten the time needed to complete the task and improved the process of gathering and analyzing the data.

**Modifications of the structure of the task.** The final version of the task contains not only the notepad where hypotheses are written but also a place where participants can make additional notes—the scratch pad. Participants claimed that it helped them to structuralize their thoughts and moved the task closer to the real-life conditions of solving a problem.

**Modifications of the instructions.** In the final version, instructions are presented step-by-step, each part immediately before the corresponding stage of the task. At first, complete instructions for the whole task overall and for each stage of the task separately were delivered at once, just at the beginning of the procedure. However, the volume of information appeared to be too extensive to be effectively processed and remembered by participants. Moreover, some participants started to perform latter stages of the task ahead of time. The content of the instructions has been refined many times so as to make them as precise as possible but with no loss of clarity. In order to assure standardization of the interaction between the subject and the researcher, a model of communication has been designed. This model applied not only to the necessary information but also to situations that may occur.

**Modifications of the procedure.** The chosen version of the procedure is set in time-restricted conditions: up to 20 min for the first stage, 30 min for the second, and 15 min for the third. These limits appeared to be a reasonable compromise between the risk of cognitive fatigue and a need to assess performance that is really a reflection of a cognitive competence in abductive reasoning, not only a mental speed or verbal fluency. What is more, the time limit seems to be closer to real-life conditions than limiting the number of questions or actions that can be performed during the task and saves the possibility of analyzing two important indicators of individual differences: the number of questions asked by the participant and the dynamics of questioning.

**Modifications of the story line to make the task easier and less formal.** When starting the second pilot study, we believed that placing the task in the context of a detective story would help to create a ludic atmosphere. As a consequence, the higher level of the participants’ engagement and, at the same time, lower level of stress associated with an experimental condition of the assignment were expected. The participants played the role of a detective seeking an explanation for what had happened in a presented enigmatic story. However, the convention of a detective story appeared to be problematic. Interviews with participants of the study revealed that some of them were misled by the background story and had searched for an explanation in a strictly directed way—they were looking for a serious crime, perceived as much more representative for this context than a traffic offence. Some other subjects reported feeling incompetent in the face of a task perceived by them as requiring specific forensic knowledge and procedures. What is more, this task condition caused emotional arousal that could influence further observations. To avoid these difficulties, some modifications to the story line were introduced. Subsequent versions of the task were set in a context, which was less connected with a professional expertise. The role to play was no longer a detective but an investigative journalist working for a news service, not a police department. However, the task was still to explain the same enigmatic situation. This informal working condition enabled us to avoid previous limitations.

**Modifications of the content of the task.** The content of the target story has been modified most frequently out of all aspects of the task. The initial version was the one used in the first pilot study, which is found in the Online Supplementary Materials (Find Our Pilot Study 1: Stories, story number 1).

Exemplary changes involved the removal of the misleading question, previously placed just after the story (i.e., “What has happened and did anyone break the law?”), clarifications and specifications of some facts from the story, adding some new information. The intention was to optimize the story as a trigger of abductive reasoning. The mystery had to be pronounced but, at the same time, more than one divergent way of solving it should be possible.

**Changes in possible answers to subjects’ questions.** At the starting point, possible responses to questions asked by participants during the second stage (called cooperation), designed to evaluate hypotheses testing phase, were as follows: yes, no, and it’s not important. Although it would be procedurally much easier and methodologically cleaner to maintain these options of reply, the trials without the third option (it’s not important) taught us a lesson: the answer it’s not important could strongly influence the direction of performing the task and impair the ecological accuracy of the measure. As the ecological accuracy was one of our most important principles, we had to prepare a broader context for the situation described in the story so as to enable researchers to answer every possible question. After the pilot study, the universe of the world described in the story was determined and described based on the data gathered during the research, reflecting answers to the most common questions asked by participants. Those who were conducting the experiment had easy online access to a document containing a description of the universe during the testing procedure so that they could reply to certain questions in the same manner. Otherwise, the standard procedure would not have been maintained, and the task could have been easier for some participants and more difficult for others. To avoid the necessity of describing each and every detail, a procedural guideline was introduced: The person who conducted the test was obliged to supplement the universe described in the aforementioned online document just after she or he had created new detail needed to answer the participant’s question. Every piece of
supplementary information needed to answer the problematic question has to be coherent with the story and cannot influence the performance of the subject too much, neither by misleading nor by facilitating adequate solution. Thanks to this move, other researchers were able to answer further questions referring to the same issue in a consistent manner.

Pilot Study Number 3: Evaluation of the Task, Exploratory Analyses

The aims of the third pilot study were to evaluate the task and explore what characteristics of the process and the products of abductive reasoning could be identified and evaluated with Find Out. As mentioned before, the designed setting should enable the assessment of adequacy and the expertise of explanations, but the intention was to go beyond them. We were therefore particularly interested in any variability of the data that could be interpreted in terms of expressions of individual differences. Twenty undergraduate students participated in the study. As very little was known about possible variables that could predict or explain any observed individual differences, an explorative approach was the one which was chosen to analyze the data.

Observations

The Find Out task enables the collection of a broad range of diversified data that can be further analyzed qualitatively and/or quantitatively. The first stage of the task—the warm-up—requires generating initial hypotheses and provides hypotheses and other notes (which cannot be coded as hypotheses). Hypotheses are preliminary proposals for explanations of what could happen in the presented enigmatic situation. They are written down in the notepad by each participant. It should be noted that in this article, we have decided to use the term explanation referring to the content of hypotheses at each stage of the task; however, more sophisticated analyses definitely suggest that not each hypothesis is an explanation in a direct sense. Other notes can be found in the scratch pad. They do not provide an explanation and can be further coded as questions, considerations, judgements, and more, for example. The second stage of the task, the cooperation, requires not only generating hypotheses but testing them as well in order to formulate the best explanation. Data from that part of the task that can be further analyzed is hypotheses (written down in the notepad), questions, justifications, and notes from the scratch pad. However, not only these products but a process can be an object of the observation as well. Exemplary procedural data that can be analyzed are the dynamics of asking questions and the modifications introduced to the content of the notepad during the phase of testing hypotheses. The third stage, the final report, consists of generating the final explanation of what exactly has happened in an unclear situation requiring an explanation. This stage gives the opportunity to analyze the final hypothesis/explanation, justifications, and other notes. What is more, there is also a possibility to compare the products of this stage with the aforementioned data, for example, the initial hypotheses or the final content of the notepad.

Samples of Qualitative Analyses

Searching for the methodological framework compatible with the primary goals of the project and enabling us to conduct exploratory qualitative research has so far directed us to the grounded theory approach and qualitative content analysis. This section of the article contains samples of the analyses that we have managed to conduct so far (some of them are still being continued). However, as a sufficiently attentive presentation of those analyses would highly exceed the acceptable limit of this article, we are currently working on a separate article presenting the results of qualitative analyses. At this point, we would only like to provide readers with a sense of the methodological possibilities of the task. Please bear in mind that the presented samples do not exhaust all the possibilities of conducting qualitative analyses of the data provided with the framework of Find Out.

The content of hypotheses: fragments and gap fillers. We conducted multilevel analyses on the formal characteristics of the content of participants’ responses. It appeared that the content of the responses could be categorized into fragments and gap fillers. Although the distinction draws on the content of the initial unclear situation, it is based not on the content itself but rather on the function of each piece of information. The initial situation is based on the narrative context, which consists of different aspects of the story, for example, police control, the breathalyzer test, and the departure of the police. Some aspects are partially further explained in the story (e.g., we know that the control was a routine one) but others are not (e.g., we do not know what the result of the test was). Each participant has to identify which information should be completed by themselves. It requires identifying the gaps—in other words, moments in the story line that are not clear enough and thus make it impossible to build a satisfactory explanation of what exactly happened in the story. Pieces of information that serve as an explanation of what exactly happened, how it happened (means, circumstances), and why (intentions, situational causes) or further specification of people, objects, locations, and time characteristics which all build the context of the story can be called gap fillers. On the other hand, fragments are basic units of information which can be easily understood even without additional information (e.g., “police stopped the car”) but can be further explained by gap fillers (e.g., when: “on a winter’s night”). However, fragments are not gap fillers themselves (because gap filler can be understood only in a relation to the fragment it is referring to). Fragments can relate to the events described in the story but also to people, objects, locations, and time characteristics of the context. It can be imagined that fragments create a sketch of the story or an explanation, whereas gap fillers complete the sketch with certain qualities. Fragments and gaps fillers can be provided in advance within content of the story or have to be determined on
Analyses revealed that basic events described in the story (understood as fragments that can be further explained in the process of filling in the gaps) can be perceived as answers to the question “What has happened?” Other fragments and gap fillers (if delivered in the content of the story or not) answer not only the aforementioned questions of “when and why” but also the questions “who, (with) what, where, what happened, and how.” It should be emphasized that fillers addressing the problem “why and how” are usually used to specify only what has happened. It can be concluded that events and fillers correspond to seven gold forensic questions (see, Gross, 1908) that should be asked by a policeman during the process of a criminal investigation: “What (has happened), who, (with) what, where, when, how, and why.”

The distinction of fragments versus gap fillers can be applied to the process of analyzing both the products (hypotheses, questions, justifications, final report) and the process of abductive reasoning (the way of modifying the content of the notepad during the process of testing hypotheses, the way of asking questions, a way of testing hypotheses, for example, identifying gap fillers and searching for information). Some examples of analyses based on that framework are presented below.

A problem-space coverage. The aforementioned distinction helped to create a coding system referring to the dimension of broadness or, in our nomenclature, a problem-space coverage. This coding system enables the researcher to assess how many aspects of the story (coded as certain fragments and initially provided fillers) someone has taken into account in the process of formulating an explanation in a hypothesis or a final report.

Analyses led us to the conclusion that the problem-space coverage did not sufficiently explain the variability of the participants’ performance regarding explanations (hypotheses and the final report), so the idea of gap fillers was further developed to a multilevel system. The first-level gap filler refers to additional information directly related to the basic elements of the story (fragments). For example—the fragment “police stopped a car” is specified by two (independent from each other) gap fillers: “on a winter’s night” (when?) and “for a routine control” (why?). In the target story, there is no other information directly connected with this fragment but the participant may offer another to develop an explanation, for example, “in accordance with police procedure” (how?). What is more, the participant can suggest additional fillers, for example, “after 8 p.m.” (when?), and it would be the second-level gap filler specifying the time of the police control (to the first-level gap filler: “on a winter’s night”). The information “in December” would be also a gap filler of the second level related to “on a winter’s night” and “on the 23rd December” would be a further explanation of when exactly the event has happened so it would be a third-level gap filler pointing the exact date (“in December”).

Two independent variables referring to the breadth and depth of explanations were operationalized. The breadth of an explanation refers to the number of alternative fillers used to fill self-perceived gaps in the story. For instance, an explanation “The driver bribed or threatened police officers to let him drive away” contains two fillers that specify the actions of the driver (he bribed or he threatened). The depth of an explanation can be operationalized as the number of consecutive fillers used to complete gaps of progressing levels of explanation. An example of a two-level explanation can be as follows: “The driver bribed police officers to let him drive away. He offered them 500 euros.” These variables can be easily analyzed quantitatively and qualitatively. An example would be to categorize individual patterns of gap filling or track how a participant changes the pattern of explaining during the second phase of the task. Gap fillers can be further categorized in many ways. As a result of our analyses, more independent categories differentiating gap fillers may be identified (e.g., regarding time orientation, the general/direct nature, or function of the gaps).

There are some perspectives of qualitative analyses that we have taken into consideration but, as both the research and analyses are still in progress, the results cannot be presented yet:

1. the correctness of the effects and the process of abducting (e.g., hypotheses, questions, justifications, modifications),
2. the manner of organizing the content of information which appeared to fit the distinction proposed by Bruner (1986, 1991) between a paradigmatic and narrative mode of cognition,
3. types of gap fillers, and
4. strategies of solving the abductive problem which can only be identified after comparing the three stages of the task (the work is still in progress).

Conclusions

The Find Out task was designed to observe and analyze not only the products but also the process of abductive reasoning in a way that takes into account both the hypotheses generation phase and evaluation phase. We believe that goal can be regarded as accomplished. Each stage of the task requires generating hypotheses and the second stage encourages generated ideas to be tested in a way preferred by the participant by asking questions. What is more, each question has to be justified, which sheds some light on the reasons for the particular decision of each participant. Similarly, the final report with the best explanation (or explanations) also contains a part with justifications. Another aspect of the process of abductive reasoning that can be evaluated is based on the dynamics assessed by comparing the results of particular stages of the task and modifications introduced by the participant during the cooperative stage.
On the basis of described pilot studies, we believe that Find Out delivers the methodological framework that enables the gathering of a broad range of data and conducting analyses of both a qualitative and quantitative nature. We are still developing the system of analyzing so as to create easier procedures of some basic assessment system that could be used by other researchers interested in the subject but not ready to conduct complex analyses.

What seems to be a reasonable doubt is whether the requirements and laboratory context of the task provide conditions that enable a process of abductive reasoning that is close enough to real-life conditions to be observed. Although our tool seems to be much closer to that requirement than other known measures of the same phenomenon, we still cannot claim that the ecological accuracy of the measure is optimal.

What can be considered an asset of the method is that it provides a framework to conduct further empirical research of both a qualitative and quantitative nature. We plan to take advantage of some possibilities and are currently working on an article presenting the results of qualitative analyses of the data. Further analyses are still being conducted in order to work out a satisfying assessment system that could serve in a future investigation focused at looking for the cognitive and personality correlates of the individual differences in performing abductive reasoning.

**Acknowledgments**

The authors would like to thank former students of the Faculty of Psychology and Cognitive Science, Dajana Bieganowska and Kinga Antonik-Jończyk, for their help during the process of conducting the research.

**Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding**

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Science Centre, Poland (DEC-2013/10/E/HS1/00172).

**ORCID iD**

Dorota Želechowska  https://orcid.org/0000-0001-6691-2350

**Supplemental Material**

The supplemental material for this article is available online.

**References**

Ajdukiewicz, K. (1974). *Logika pragmatyczna* [Pragmatic logic]. PWN.
Bruner, J. (1986). *Actual minds, possible worlds*. Harvard University Press.
Bruner, J. (1991). The narrative construction of reality. *Critical Inquiry, 18*(1), 1–21. https://doi.org/10.1086/448619
Donnelly, M. B., Sisson, J. C., & Woolliscroft, J. O. (1990). The reliability of a hypothesis generation and testing task. *Medical Education, 24*(6), 507–511. https://doi.org/10.1111/j.1365-2923.1990.tb02666.x
Gross, H. (1908). *Handbuch für untersuchungsrichter als system der kriminalistik* [Handbook for examining magistrates as a system of criminalistics]. Leuschner & Lubensky’s.
Hobbs, J. R. (2006). *Abduction in natural language understanding.* In L. R. Horn & G. L. Ward (Eds.), *The handbook of pragmatics* (pp. 729–741). Blackwell. https://doi.org/10.1002/9780470756959.ch32
Ingram, C., Ray, K., Landeen, J., & Keane, D. R. (1998). Evaluation of an educational game for health sciences students. *Journal of Nursing Education, 37*(6), 240–246. https://doi.org/10.3928/0148-4834-19980901-04
Kwon, Y. J., Jeong, J. S., & Park, Y. B. (2006). Roles of abductive reasoning and prior belief in children’s generation of hypotheses about pendulum motion. *Science & Education, 15*(6), 643–656. https://doi.org/10.1007/s11191-004-6407-x
Kwon, Y. J., Lee, J. K., Shin, D. H., & Jeong, J. S. (2009). Changes in brain activation induced by the training of hypothesis generation skills: An fMRI study. *Brain and Cognition, 69*(2), 391–397. https://doi.org/10.1016/j.bandc.2008.08.032
Lawson, A. E. (1978). The development and validation of a classroom test of formal reasoning. *Journal of Research in Science Teaching, 15*(1), 11–24. https://doi.org/10.1020/tea.3660150103
Lawson, A. E., Clark, B., Cramer-Meldrum, E., Falconer, K. A., Sequist, J. M., & Kwon, Y. J. (2000). Development of scientific reasoning in college biology: Do two levels of general hypothesis-testing skills exist? *Journal of Research in Science Teaching, 37*(1), 81–101. https://doi.org/10.1002/(SICI)1098-2736(200001)37:1<81::AID-TEA6>3.0.CO;2-I
Lee, J. K. (2012). Do the brain networks of scientists account for their superiority in hypothesis-generating? *International Journal of Environmental and Science Education, 7*(1), 83–106.
Magnani, L. (2009). Abductive cognition. *The epistemological and eco-cognitive dimensions of hypothetical reasoning*. Springer.
Mirza, N. A. (2015). Effects of abductive reasoning training on hypothesis generation abilities of first and second year baccalaureate nursing students (PhD thesis). McMaster University.
Oh, P. S. (2008). Adopting the abductive inquiry model (AIM) into undergraduate earth science laboratories. In I. V. Eriksson (Ed.), *Science education in the 21st century* (pp. 263–277). Nova.
Ottens, A. J., Shank, G. D., & Long, R. J. (1995). The role of abductive logic in understanding and using advanced empathy. *Counselor Education and Supervision, 34*(3), 199–211. https://doi.org/10.1002/j.1556-6978.1995.tb00242.x
Peirce, C. S (1931–1935) Collected papers of Charles Sanders Peirce (Vols. 1–6; C. Hartshorne & P. Weiss, Eds. Vols. 7–8, 1958; A. W. Burks, Ed.). Harvard University Press.
Russo, J. E., & Meloy, M. G. (2002). *Hypothesis generation and testing in Wasons 2-4-6 task.* citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.569.3425&rep=rep1&type=pdf
Schurz, G. (2017). Patterns of abductive inference. In L. Magnani & T. Bertolotti (Eds.), *Springer handbook of model-based science* (pp. 151–173). Springer. https://doi.org/10.1007/978-3-319-30526-4_7
Tavory, I., & Timmermans, S. (2014). *Abductive analysis: Theorizing qualitative research*. University of Chicago Press.

Thagard, P. (2000). *Coherence in thought and action*. MIT Press.

Thagard, P., & Shelley, C. P. (1997). Abductive reasoning: Logic, visual thinking and coherence. In M. L. Dalla Chiara, K. Doets, D. Mundici, & J. van Benthem (Eds.), *Logic and scientific methods* (pp. 413–427). Kluwer AP.

Urbaniński, M., & Klawiter, A. (2018). Abduction: Some conceptual issues. *Logic and Logical Philosophy*, 27(4), 583–597. https://doi.org/10.12775/LLP.2018.016

Vertue, F. M., & Haig, B. D. (2008). An abductive perspective on clinical reasoning and case formulation. *Journal of Clinical Psychology, 49*, 1046–1068. https://doi.org/10.1002/jclp.20504

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology, 12*, 129–140. https://doi.org/10.1080/17470216008416717.

Woods, J. (2017). Reorienting the logic of abduction. In L. Magnani & T. Bertolotti (Eds.), *Springer handbook of model-based science*. Springer. https://doi.org/10.1007/978-3-319-30526-4_6