Reproducibility and Instruction Following in the Shop Floor Laboratory Work: The Case of a TMS Experiment

Kristina Popova

Abstract
The article addresses the production of reproducibility as a topic that has become acutely relevant in the recent discussions on the replication crisis in science. It brings the ethnomethodological stance on reproducibility into the discussions, claiming that reproducibility is necessarily produced locally, on the shop floor, with methodological guidelines serving as references to already established practices rather than their origins. The article refers to this argument empirically, analyzing how a group of novice neuroscientists performs a series of measurements in a transcranial magnetic stimulation experiment. Based on ethnography and video analysis, the article traces a history of the local measurement procedure invented by the researchers in order to overcome the experimental uncertainty. The article aims to

1KTH Royal Institute of Technology, Division of Media Technology and Interaction Design, Stockholm, Sweden

Corresponding Author:
Kristina Popova, KTH Royal Institute of Technology, Division of Media Technology and Interaction Design, Lindstedtsvägen 3, 114 28 Stockholm, Sweden.
Email: kpopova@kth.se
demonstrate (1) how reproducibility of the local procedure is achieved in the shop floor work of the practitioners and (2) how the procedure becomes normalized and questioned as incorrect in the course of experimental practice. It concludes that the difference between guidelines and practical actions is not problematic per se; what may be problematic is that researchers can be engaged in different working projects described by the same instruction.

**Keywords**
cognitive neuroscience, replication crisis, instruction following, reproducibility, ethnomethodology, video analysis

**Introduction**

The crisis of replication has been recently proclaimed across multiple scientific fields (Button et al. 2013; Open Science Collaboration 2015; Stevens 2017), drawing attention to the question of how reproducibility is produced. The crisis has highlighted the “backstage” of scientific work (Freese and Peterson 2018)—seemingly insignificant research practices that were not earlier relevant for anybody except scientists themselves and the early lab ethnographies (Latour and Woolgar 1979; Knorr-Cetina 1981; Lynch 1985). Several recent studies have critically addressed the contrast between clear descriptions of methodology and the contingencies of daily research work, implying that the difference between those two can possibly confound research results (Leahey 2008; Simmons, Nelson, and Simonsohn 2011; Peterson 2016). A related branch of research has explored the pervasiveness of “questionable research practices” located in the “gray areas” of scientific ethics—not frauds but actions potentially violating “the standards of good science” and hence impairing the reproducibility of scientific results (John, Loewenstein, and Prelec 2012; Fiedler and Schwarz 2016).

The problem of replication has been familiar to the practicing scientists and scholars of science long before this crisis. Replication has been problematic in reproducing new results on the cutting edge of science (Collins 1989, 2001), transforming a relatively well-established technique into new contexts (Jordan and Lynch 1998), educating new generations of scientists (Amerine and Bilmes 1988; Roth and Bowen 2001), and even repeating the same procedure within the same research team (Roth 2009). What is new in the recent discussions is that the topic of replication has been critically
addressed from the outside, and the lack of reproducibility has been marked as a problem that could potentially be solved by eliminating a gap between research practices and their methodological descriptions.

The gap between reported research methodologies and shop floor practices is problematic when the idea of reproducibility is based on the premise that a published research report should allow other scientists to reproduce the initial study (Goodman, Fanelli, and Ioannidis 2016). From this perspective, both deviation from guidelines and failure to report all the details of undertaken studies are harmful for the scientific community. A number of solutions have been proposed, including appeals to follow the standards more thoroughly (Leahey 2008) and to disclose more information on research methodology in the reports (Simmons, Nelson, and Simonsohn 2011; Stevens 2017). It is noticed, however, that the level of reported details required to ensure methodological reproducibility is highly ambiguous (Goodman, Fanelli, and Ioannidis 2016, 2), and even when the research methodology is thoroughly described, it does not eliminate all the possible ambiguities embedded in the ostensibly simplest steps (Simmons, Nelson, and Simonsohn 2011, 1360). Attempts to estimate the scale of violations have also faced the problem of classification upon discovering that whether a research practice is a questionable, justifiable, or normal depends on the conditions of its application. As reported in a study of questionable research practices: “Although falsifying data [...] is never justified, the same cannot be said for all of the items on our survey; for example, failing to report all of a study’s dependent measures [...] could be appropriate if two measures of the same construct show the same significant pattern of results but cannot be easily combined into one measure” (John, Loewenstein, and Prelec 2012, 531). Their discussants emphasize similar problems, arguing that, for instance, selective reporting of positive results is not necessarily questionable as it may refer “to completely normal practices that are part and parcel of all empirical research,” such as conducting pilot studies or testing experimental conditions (Fiedler and Schwarz 2016, 46). The reproducibility crisis therefore invokes the old theoretical paradox of rule-following, which rests on the fact that “there are no rules on how to apply rules” (Ogien 2009, 461) and, consequently, rules to define a correct application of the rule.

Laboratory studies have addressed the gap between written instructions and practical actions through references to tacit knowledge—expert competence that allows scientists to take decisions in the absence of any formalized criteria (Collins 1989, 2001; Polanyi 1983). Outside science, expert judgment, as an ability to creatively apply rules in practical situations
based on one’s previous experience, has been recognized as a general feature of a skillful performance (Dreyfus, Dreyfus 1980). Drawing attention to the fact that the problematic gap between rules and situated actions rarely disappears completely, Roth (2009) has used the concept of radical uncertainty to describe the problems that experienced scientists, who already possess tacit knowledge, face while following familiar methodological procedures and yet still encountering the gap between plans, situated actions, and their methodological descriptions (Suchman 1987).

The present article aims to analyze how uncertainty is addressed in the work of novice researchers performing an experiment in cognitive neuroscience with the method of transcranial magnetic stimulation (TMS). The empirical part explicates how novice experimenters approach uncertainty in the measurement of a motor threshold (MT) by developing a specific measurement procedure grounded in the requirements of their experiment, which they later have to justify to another laboratory member. Based on ethnography and video analysis, the paper address two issues: (1) how reproducibility of the local measurement procedure is established in situ, that is how the procedure becomes reenacted on different occasions and recognizable for both the participants and analysts as different occasions of the same method; and (2) how the scientific credibility of the procedure is negotiated, established, and contested within the group of experimenters and in communication with another laboratory member.

Drawing on ethnomethodological studies of rule following and scientific practice (Amerine and Bilmes 1988; Alac 2011; Bjelic and Lynch 1992; Garfinkel 1967, 2002; Garfinkel, Lynch, and Livingston 1981; Lynch, Livingston, and Garfinkel 1983; Sormani 2014), this article is based on the premise that reproducibility of practices is necessarily produced in situ, in the shop floor work of practitioners, with guidelines serving as references to the already established ways of work. As distinguished from the earlier studies of uncertainty in the work of novice researchers (Delamont and Atkinson 2001; Goodwin 1994, 2000) and ethnomethodological studies of science, this paper focuses on the local history of the measurement procedure developed by the group of experimenters.

Since the procedure becomes invented due to the experimental requirements, I first describe these requirements along with a guideline for finding an MT. After that, the transcripts of four chronologically recorded video fragments are presented to illustrate the establishment and abandoning of the local procedure. The first three fragments demonstrate how the procedure becomes introduced, established, and taken for granted as a part of experimental practice of the group. In the fourth and final fragment, the
procedure is contested by another lab member and has to be explicated again in order to be seen as an established approach rather than a mistake. Although the procedure is acknowledged to be a systematic method of measurement, it is still qualified as a mistake by the laboratory member who does not share the criteria justifying the relevance of the “in-between” method within a particular experiment. The analysis of the last fragment demonstrates that the procedure is evaluated as part of a working project that determines whether the deviations from and adjustments to a guideline are insignificant or crucial enough to become an error.

The comparison between guidelines and observable actions does not help to identify a mistake since the total match between practical actions and their “formal descriptions” (Garfinkel and Liberman 2007) is not achievable. Pointing to the mismatch between instructions and situated actions may be a strategy used by external critics (Lynch 2002), but it is not something problematic on its own for practitioners. An alternative project may involve analyzing the criteria that practitioners apply to distinguish between correct and incorrect procedures and establishing whether practitioners are involved in the projects producing the same identification of mistakes.

**Data and Settings**

The data analyzed in the empirical part had been collected during two-year video ethnographic project conducted in the laboratory of TMS. Using ethnomethodologically based video analysis (cf. Heath, Hindmarsh, and Luff 2010) as a main method, the paper follows the experiment of a research group of two novice researchers who are postgraduate students working in the lab to collect data for their research project. By the time the first analyzed fragment was filmed, the researchers had spent approximately three months in the lab, performing a pilot study and conducting training sessions before launching the main data collection project. The experimental practice of the research group is hence unsettled enough to let us notice changes and aspects that may be taken for granted in the work of more experienced researchers. The fact that the experimenters are students may cast doubts upon the credibility of the analysis, as students can be easily criticized for not being “true scientists.” It should be noticed in that regard that distribution of duties in many labs, not excluding this one, presupposes that responsibility for conducting full range of research activities, including data collection and reporting results in publications, rests with postgraduate students, especially doctoral students (Peterson 2016), while senior lab
members are mainly engaged in administrative work, writing, and preparing grant proposals.

The analysis is built on four video records chosen from the longitudinal data set consisting of approximately twenty hours of video data that chronologically depict work of several experimental groups and singular experimenters. The fragments are transcribed using a modified Gail Jefferson’s system, whose conventions are described in Appendix. Field notes and excerpts from the semi-structured interviews are used as supplementary materials. The transcribed conversations are originally in English (the participants are second-language speakers), and the interviews were translated. The choice of the fragments mainly stems from the purpose to demonstrate long-term development of a procedure. In addition, the procedure is grounded in commonsense mathematical reasoning and hence is comprehensible regardless of the reader’s background. Last but not least, it is explicated through the interaction between the researchers, which makes it more accessible for the analyst.

While measuring MT is crucially important for the validity of TMS experiments (Shutter and Honk 2006), the analyzed case demonstrates the contingency of laboratory work in “safe settings.” Variability of TMS responses is high enough to make most of the experimenters’ innovations insignificant compared to other sources of MT variations (Shutter and Honk 2006; Sommer et al. 2002; Tranulis et al. 2006). At the same time, as the nature of TMS variability is still understudied, the analysis may point to another possible source of MT variability, that is, the difference in the measurement practices.

**TMS**

To put it simply, TMS works the following way: a magnetic coil (see Figure 1) located on a particular zone of the brain cortex produces an electromagnetic field, which evokes motor reactions or affects participants’ cognitive functions (Rossi et al. 2009). Depending on the coil’s position, which is controlled by experimenters, stimulation can affect participant’s visual system, language production, or motor functions (Hallett 2000, 147). The changes in motor functions are detected by electrodes, which are placed on participant’s muscles to measure muscle contraction as motor-evoked potential (MEP), which is displayed on one of the monitors in the lab. The experimenters, therefore, can detect some results of stimulation in real time by observing the fluctuation of the MEP curve on the monitor (some of the effects can also be observed visually, for instance, through the movement of
participant’s fingers). This is the type of data that experimenters need for finding MT.

MT is measured in order to normalize the intensity of stimulation received by each participant. Since all participants have different sensitivity to magnetic stimulation, giving all of them the same amount of stimulation will not produce comparable results. The experimental intensity of stimulation is thus usually counted in relation to the minimal response of each participant, which is called MT. MT is a measurement needed to control for interpersonal differences in reactions to TMS. MT is formally defined as a “minimum TMS intensity sufficient to produce a predefined motor-evoked potential (MEP) in the contralateral abductor pollicis brevis in at least 50% of trials” (Herbsman et al. 2009, 3, citing Rossini et al. 1994). Finding an MT, therefore, is crucial to ensure that during the experimental session all the participants are getting equal amounts of stimulation, adjusted for their sensitivity.

In the studied TMS lab, there is a guideline prescribing the necessary steps for finding MT. The guideline presented below is taught in the training
sessions conducted in the lab and, with some variations, appears in the research papers produced by the lab members. The following description, therefore, is a reconstruction of the algorithm that the experimenters are taught and later apply in the studied lab. It coincides with the description provided for the Rossini–Rothwell algorithm—one of the three most widespread methods for MT search.

**The Lab’s Guidelines for Finding MT**

1. Identify a hot spot, that is, a place on the brain motor cortex wherein TMS evokes the muscle contraction;
2. Find a minimal intensity which evokes the response exceeding fifty hertz in five out of ten pulses.
   a. Stimulate the spot ten times and count the number of sufficiently high responses.
   b. Decrease the intensity if there are more than five responses; increase it if there are less than five.
   c. Repeat.
3. Check the responses on neighboring spots, making sure that there is no spot with a better response on lower intensity.
4. Repeat previous steps if necessary.

For one who is not competent in TMS, the guideline contains a number of blind spots. Most noticeably, when understood strictly according to the guideline, the third step of checking other potential spots can be performed endlessly, since there are no formal criteria for when a researcher should stop checking whether there is another spot that would give responses on lower intensities. This is a moment of uncertainty, where experimenters have to express their expert judgment and decide when it is appropriate to stop searching. This is also a moment when most of the local disagreements around the correctness of the experimenter’s decision emerge.

Measurement of MT is highly problematic for novices, who have to distribute their attention between multiple details of the procedure (position of a magnetic coil, location of electrodes, display of MEP signal) and perform the decision-making under uncertainty. The novices have also to keep in mind that the procedure is variable, and the participant’s motor responses may change not only due to the experimenters’ actions but also because of other reasons, as the instability of motor responses is acknowledged to be one of the characteristics of the TMS (Darling and Butler 2006). It means that experimenters should distinguish between changes evoked by
their actions, such as rotation of a magnetic coil, and the changes occurring due to the inherent instability of the procedure or slight inaccuracy of the measurement tools. The citation from the interview with one of the experimenters demonstrates that the instability is problematic for those who are trying to master the procedure:

This [searching for MT] is a kind of slightly variable thing. We can, for instance, find an okay point for the moment, then increase intensity a bit—and that’s it—we don’t see responses anymore. It is important to understand that, first, everything is slightly bouncing and you must orient in the dynamic environment, because those personal maps they change sometimes. I’m exaggerating a little now, but it’s really not a super stable thing, and it demands certain confidence. [. . . ] Over time you just learn to correctly distribute the distance between spots and choose intensity to find the area of interest, but it requires confidence. It’s possible to start doubting on each step, there are many decisions taken on each of them and those decisions require certain boldness. (Interview with a novice researcher working in the lab)

As it follows from the quotation, experimenters have to develop both systematic procedures, allowing one to choose, for instance, the distance between spots or intensity of stimulation, and the “boldness” to stop at the decision as being good enough. The “boldness” highlighted in the quote aligns with what ethnomethodologists refer to as “ad hoc” considerations—situated concerns allowing one to set correspondence between a guideline and the observable activity (Garfinkel 1967, 20-21). In the following sections, development of a systematic way to handle the uncertainties of MT search is addressed.

**The Relevancies of the “In-between” Procedure**

The specific ambiguity in searching for MT consists in the requirement to have sufficient responses “in 50 percent” of stimulation attempts. The practical approximation of this rule implies that, after discovering a “hot spot” with visible motor responses, an experimenter gives ten impulses and calculates the number of attempts that produced MEP above a certain number. “Fifty percent” then means five successful responses. If the number of successful responses exceeds five, the stimulation intensity should be reduced; if it is fewer than five, the intensity should be increased. The problem emerges when experimenters have either more or fewer responses on the adjacent intensities of stimulation (four or six instead of five), which, according to the procedural guideline taught in the lab, is too low or too high
to establish a threshold. Experimenters then face a dilemma, whether they should take one of the tried intensities as a threshold, try again, or check another spot. The possible instability of TMS responses, the conventional character of the procedures (there is no clear explanation why one is supposed to try the point ten times, rather than eight or twelve) and time limitations complicate the process even further.

In the analyzed experiment, additional complications emerge because experimenters have specific requirements for the spot that they have chosen as a “hot spot” to measure MT. After finding MT, they will need to stay at this hot spot and perform several other steps to record data. Due to the nature of the next steps, the spot needs to evoke a proper balance of responses in two measured muscles—first dorsal interosseous and abductor digiti minimi. The specifics of the data collection phase, therefore, place additional requirements on the responses obtained during the measurement of MT. Since not all the spots are suitable to continue the experiment, repositioning the magnetic coil is more time-consuming for these experimenters than for other researchers measuring MT.

Based on the specifics of the experimental requirements, the experimenters developed a simple solution to the problem of having more or less than five responses at the adjacent intensities at a “good enough” spot. The decision consisted in staying at the spot and taking the average intensity as a threshold. For example, if an experimenter finds a hot spot with the suitable balance of muscle responses but gets slightly less or slightly more responses to call it a threshold, she takes an intermediate value as a threshold. If a researcher gets six of the ten pulses on the 51 percent of intensity and then four of the ten pulses on 50 percent, she may stop searching and take an average score (50.5 percent) as a threshold. The choice of a nonintegral MT is not problematic, since MT serves as an intermediate value, which is only needed to calculate the intensity of experimental stimulation moving to the next step.

As was explained by the researcher who introduced this procedure, the decision relates to the variability of TMS responses and makes MT search more precise and less arbitrary:

Well, it [taking an average intensity as a threshold] relates to the situation with variability once again. I am pretty sure that if I took intensity 44 and make ten more pulses, there could be seven or four responses, not five. It is optional averaging, but it seems logical to me. If you have four on forty-three and six on forty-four, by logic, the threshold is an intermediate value somewhere between. Obviously, it is an intermediate value. Since we are checking 120% from this value, we can be sure that this is the right intensity.
The “in-between” procedure is one of the several modifications that the studied group introduced in the MT measurement procedure to adapt it to their practice. Reducing the number of stimulations and spots tested after finding MT are among other small variations. The “in-between” procedure is the most systematic adjustment that substantially transformed how the group perceives MEP sensors and their actions. The following section demonstrates how the in-between procedure has first become a part of experimental practice and was later questioned by another lab member.

**The Local History of “In-between” Procedure**

This section analyses the development of an “in-between” procedure to demonstrate how it becomes established and reproducible within the specific data collection project, and how it is later dismissed by another lab member as a mistake. The history of the procedure is used to discuss the relationships between guidelines, reproducibility, and mistakes in the concluding section.

**The First Appearance**

P -- Petr  
T -- Tanya²

1. P  (gives a pulse) (1.6)
2. actually I would say that this is kind of a thres[hold
3. T  [yeah=
4. P  >maybe something between< we can take (0.7) fifty: (0.5) five (1.0)
5. (hangs the coil)
6. a:: nd (. ) transform (0.4) transform it (1.6)
7. (steps away from the chair)
8. <*>to one hundred and twenty percents*> (0.9)
9. T  (takes the pen, looks at stimulator, then at P)
10. this is fifty ( . ) if we take fifty one?
11. (1.7)
12. P  (looks at T)
13. le: t’s just check ((unclear))
14. (walks away)
For the first time, the procedure is introduced by Petr, the experimenter responsible for finding MT, and exposed to his colleague Tanya who is supposed to approve the threshold and write it down in the laboratory journal (Figure 2). The decision to take an average intensity is initially proposed as a test. Petr begins by expressing uncertainty: “I would say,” this is “a kind of a threshold,” “maybe something between” (lines 2 and 4). This uncertainty, however, does not invite further negotiations, since he hangs down the coil and moves away from the chair before his colleague agrees with the decision. He also does not provide any clarifications when Tanya does not demonstrate understanding and asks a question in line 10. The procedure needs to be approved before it becomes established, but the criteria of approval pertain to practice; it has to be tested by further experimental steps.

The background rationales standing behind the procedure are not discussed explicitly. Although the procedure is built on the relevancies significant for this experiment, those relevancies are taken for granted. However, Petr outlines the place of a new procedure in the experiment. He first names the results, that is, an average threshold intensity that the experimenters will need on the next experimental step (fifty point five), and then points at the next step to undertake, that is, calculating 120 percent of the threshold intensity. No other accounts of a new procedure, except that it can be integrated into the data collection process, are provided during its first use. Here, the question of whether the procedure is correct relates to the question of whether it can be included in the main task of data collection rather than to the correspondence with the general guideline. “Let’s just check” serves as a sufficient justification at this stage since the main criteria of methodological credibility pertain to

**Figure 2.** Searching for a motor threshold.
practice. Taking “in between” numbers is yet to be established as a part of the experimenters’ practice, but the work of establishing begins only when the procedure passes the test successfully, producing the results suitable for the continuation of the experiment.

The Explication

The second use of the procedure occurred several weeks later when the group once again faced a problem of having either more or less than half of the sufficient responses on the neighboring intensities of stimulation. At that point, the procedure had been tested by subsequent experimental steps, and so it could become a part of the experimenters’ repertoire. However, it had not been yet established within the research group. As the lack of understanding was demonstrated previous time, the researchers are engaged now in establishing the procedure.

1. P (looks at navigator, navigates the coil)
2. (looks at the recorder, gives a pulse)
3. [(looks at navigator)
4. [hm:
5. (1.3)
6. (gazes at T)
7. (looks at navigator, raises the coil)
8. it seems like it’s a bit too [low
9. [(reaches his pocket with a phone, looks at T)
10. (0.5)
11. T (turns to P)
12. it’s in between so=
13. P [(takes out the phone)
14. ye[a:: h (.) let’s [try calculate on
15. [(calculates on his phone)
16. T [(turns to the journal, takes a pen)
17. (2.3)
18. P from forty six and five and the: n (0.9) *see if it’s okay*
19. T [(writes in the journal)

The conversation is again initiated by Petr who is searching for MT. When eight pulses have been given, he interrupts the stimulation, checks the monitors
where the coil’s position and MEP are displayed, and produces a “hm,” marking that there is something to be noticed. He then briefly gazes at his colleague, lifts the coil, and sets up the situation: “It seems like it’s a bit too low.” Here, Petr acts as a more knowledgeable participant who is making sure that Tanya notices the same details leading to the use of the procedure as he does. Since the number of responses is “a bit” low, rather than just low or too low, they take an average value instead of checking other spots or intensities.

The description is followed by a 0.5 second pause when Petr is reaching for his phone, which he will need on the next step to calculate 120 percent from the threshold, and looks at his colleague who eventually accepts his assessment and proclaims that the threshold is “in between.” By doing so, she establishes the name of the procedure building on Petr’s description of the situation. Her choice of naming “in between” demonstrates the understanding of similarities between two cases as it refers to the previous experimental session, where the procedure was described as “something between.” The naming is one of the “substrates” (Goodwin 2018) that provides continuity between two interactional moments ensuring that the procedure is observably reproduced as a new iteration of the same method.

This demonstration of understanding and the choice of procedure are then acknowledged by Petr’s “yeah” in line 14. Together with this acknowledgment, Petr starts calculating and simultaneously commenting on the result of measurement and his actions (“let’s try calculate from forty-six and five and then see if it’s okay”). As a response, Tanya turns back to the journal to finish her part of the procedure and writes down the threshold number, making visible that the procedure has become integrated into the group’s experimental practice.

During the second usage, the constitutive steps of the procedure are exposed in the cooperative work, which starts from Petr’s definition of the situation, followed by Tanya’s naming of the procedure, and Petr’s proclamation of measurement results and further steps (calculating 120 percent). As we shall see in the following excerpts, those elements are reproduced on the different occasions where the procedure is applied, building the local accountability of the procedure as a systematic way of measurement and making the cases of its application recognizable as the same but new situations both for us as analysts and the participants. The explication of the method accomplishes several tasks: establishing the procedure as an acceptable way of measurement within the team, ensuring the understanding between the experimenters, and serving as a pedagogical device that allows Tanya to master the procedure.

However, this work is embedded in the main task of experimental data collection and becomes a subproduct of the main practical task of producing
experimental data. As well as in the previous fragment, the researchers are mainly engaged in collecting or fixing information needed on the following steps of the experimental procedure. The talk serves as a supplement to the main course of action. In a similar way, understanding and mastering the procedure is integrated in the process of data collection. As has been shown in the previous fragment, the method does not emerge as a recognizable one from the outset; its recognizability has to be achieved. However, it is primarily achieved by integrating the method in the data collection process, while instructions and negotiations play a subordinate role to the practical actions of experimenters.

**Becoming Taken for Granted**

In the second fragment, the elements of the “in-between” procedure, including the definition of situation, its name, results, and the following step, were explicated to make them recognizable for the participants. During the next reiteration, which was recorded another week later, those elements start to be taken for granted as the experimenters can support mutual understanding without speaking through the steps of the procedure.

1. P  (navigates the coil, gives a pulse)
2.   hm:  
3.   (0.7)
4. T  (turns to P, looks at him)
5. P  [(raises the coil, steps away)
6.   [Well, actually I would take: a: [thirty three point five
7.    [thirty three point
8.     [five (0.3)
9. P  yes=
10. T  =mhm
11.   (starts calculating)
12. P  (hangs down the coil, picks up the phone, starts calculating)

The structure of the situation, as it is depicted in the transcript, remains very close to the previous excerpt. Petr starts from the same “hm” sound, although now his colleague immediately responds by turning around and making eye contact with him. The description of the situation is now
omitted; the experimenters proceed directly to announcing MT, which Petr begins from the familiar preamble (“well, actually,” “I would take”) and the set of movements (raising the coil and taking a step away from the chair). When he hesitates a bit, making a lengthy “a” sound, Tanya quickly continues the announcement of an average threshold intensity, which the experimenters name together. The team members additionally confirm mutual understanding by nodding and saying an affirmative “yes” and “mhm” in lines 8–10. Both then immediately engage in the next procedural steps, starting to calculate and writing down the threshold intensity without further discussions.

The patterns of the procedure and the conditions of its application are now recognizable for both participants, so they can skip the description of situation, naming the procedure or outlining further steps. Despite the fact that most of the elements are omitted, it does not provoke confusion within the team. The procedure is reproduced even if its elements have become taken for granted, as it became established and recognizable. The smooth and uninterrupted sequence of actions demonstrates that the procedure has been embedded into the experimental practice of this research group, and no additional efforts are required to justify its applicability or its (by now) tacit, structure. The procedure has become a part of the experimenters’ tacit knowledge, even though it can be explicated again, as happens in the next fragment, when the procedure is contested by a lab member not sharing the knowledge of the group.

**Being Contested**

Together with becoming established the structure of the measurement procedure has also become taken for granted. In the last fragment, it has to be explicated again for an outsider, laboratory assistant Marina, who is experienced in TMS but not aware of the details of this group’s experiment. Marina used to help this research team mastering TMS, although by now she is expected to assist Tanya in the absence of Petr rather than teach or supervise her. Nevertheless, she opposes the procedure and eventually persuades Tanya to follow the “regular algorithm,” which results in finding lower threshold intensity and is treated as a demonstration of the “in-between” method’s inaccuracy.

1. **T** Yeah this is [(speaks unclear)]
2. **M** [one more (.)] just do one more
3. (3.3)
4. T (puts the coil on target, makes one more stimulation)
5. M yeah. *okay* (0.9)
6. (nods, turns to T.)
7. T (raises the coil, steps away from the chair)
8. so we can take (0.8)
9. M or you can [can
10. T [forty three point (.)] fi: ve maybe:: (.)
    because it’s something
11. in between (0.4) [forty four is really high=
12. M [but you cannot put]
13. =you cannot put points
14. (0.4)
15. T (reaches her pocket, starts to take out the phone)
16. yeah, but then we (.). calculate (.). one hundred twenty::
17. (0.2)
18. M a:::. (.). bu: t (0.8) >do you want maybe< just to
    search around?
19. maybe it is not (.). [this (axis) (0.2)
20. T [(continues calculating)
21. M or you always do it like this?
22. (0.3)
23. T we always do it [like this]
24. M [a::: (shrugs, nods)] okay=
25. T =>I mean I< I can (.). search (.). [a bit around
26. ((takes out the phone)
27. (1.2)
28. (hides the phone, raises the coil)
29. T but here (.). we have really good FDI::: and so (0.2)
   (*may[be]*)
30. M [okay
31. (1.0)
32. T I will try three points around
33. M (turns to monitors)

This time, Tanya is searching for MT, while Marina helps her to monitor responses and manage the navigation interface. After Tanya gave nine stimulations, she started to produce the decision. However, she was interrupted by Marina who requests to give the tenth pulse as it is required by the standard procedure of a threshold search. The sequence of correction—
repair—evaluation featured in the first six lines depicts the contention over the appropriateness of the experimenter’s actions. The reduced number of pulses was repeatedly used by the group before, but an “outside” observer—the one who is not quite familiar with the group’s experimental practice—sees it as a mistake.

The confrontation repeats when, after giving an additional pulse, Tanya attempts to announce the decision again. She starts from the proposition (“so we can take . . .”), which is again challenged by Marina who tries to propose an alternative strategy and contests the idea of taking a decision right now (“or you can . . .”). At this point, the definition of the situation (where threshold decision can be taken) is not shared. The experimenter is ready to take the decision because the balance of muscles responses is good, and the situation resembles previous instances when the “in-between” method was used, but these details are not salient for the outsider.

This attempt to correct, however, is interrupted by Tanya who continues an MT announcement and, using the by now familiar procedure, proposes to take an average threshold value. As in the previous cases, her announcement is made as a proposal marked by “maybe” and “we can take.” She starts directly from the results of measurement (“forty-three point five”), but when no understanding or approval is demonstrated, she provides additional accounts referring to the name of the established procedure (“because it’s something in between”) and describing the situation (“forty-four is really high”), which justify its use.

These elements are still not enough to make the experimenter’s actions recognizable as a procedure, and the decision is again contested. The next correction (“you cannot put points,” i.e., use fractional numbers as thresholds) also relates to the difference between the standard procedure and the one that Tanya attempts to use. As a response, Tanya outlines the place of the procedure in the data collection project by saying that they can later calculate “one hundred twenty” percent from the threshold. She therefore explicates all the elements of the procedure that had been established in the previous sessions (definition of situation, name of the procedure, measurement results, and implications) while trying to make her way of measuring MT visible as a procedure and a part of the existing practice. After that explication, Marina changes her corrections to milder suggestions and, eventually, acknowledges the procedure as an established way of measurement within the group, who “always do it like this.” Through the experimenter’s guidance, the procedure becomes visible as the group’s method, with the conditions of its application noticeable and its steps recognizable as parts of the procedure. It leads to a mitigation of the criticism and prompts
Marina to accept the divergence from the standard procedure in lines 24 and 30. One of the reasons for temporary acceptance of a new procedure is that maintaining the uniformity of experimental practice is seen as one of the criteria for scientific rigor in the lab. Thus, even if the new addition of an experimental procedure is not considered trustable, per se, the laboratory assistant may think that uniformity of experimental practice is more important than some minor changes.

At the same time, the acknowledgment of the “in-between” measurement as a procedure does not necessarily make it correct. At the end of the transcribed excerpt, Tanya agrees to make a compromise and continue searching for MT a bit longer. Several minutes later, she finds a different spot with a sustainably lower response and establishes a lower MT. Both the experimenter and lab assistant consider the finding a demonstration of the inaccuracy of the experimenter’s approach. The acknowledgment comes in the ironically overemphasized, yet clearly seen as a criticism, form: “Just imagine if you stayed at that point! It would be forty…how much, forty-three forty-four coming at the end!” The “in-between” method and the experimenter’s strategy appear to be mistakes rather than optimization, when the results of measurement point out that the experimenter’s strategy was not providing good enough (the lowest possible) results.

The identification of a mistake results from a constellation of circumstances. The laboratory assistant is considered to be more knowledgeable than the experimenters; the experimenter lacks confidence in applying the “in-between” method, which may bring hastiness into her decision-making. The procedure becomes questioned because the lab assistant is an outsider who does not share the “external” criteria of defining the threshold specific to the group’s experiment. The lab assistant’s concerns related to the deviations between guidelines and the experimenter’s actions open up the space of uncertainty. However, the experimenter’s strategy gets dismissed as a mistake only when a lower threshold was discovered. As in the introduction of the “in-between” method, the criteria of methodological correctness pertain to practice that turns a potentially reasonable adjustment into a mistake.

Discussion and the Concluding Remarks

Previous studies of laboratory practice (Collins 1989; Garfinkel, Lynch, and Livingston 1981; Lynch 1993; Lynch, Livingston, and Garfinkel 1983) have demonstrated that total match between methodological descriptions and the implemented actions is unreachable, and the critical comparison will inevitably reveal some misalignment as the instructions never fully describe the
practice—there is always a possibility to find a discrepancy by zooming in the
details of the process. Pointing to the difference between protocols and actual
practices is a common form of external criticism, coming from administrators
or rivalry laboratories (Collins 1989; Jordan and Lynch 1998; Lynch 2002)
rather than a shop floor strategy of resolving uncertainty. To analyze how the
border between methodologically correct and incorrect measurement proce-
dure is drawn in practice, this article approached the local controversy arising
between the researchers conducting a TMS experiment and the laboratory
member not engaged in the group’s experiment.

The empirical part has demonstrated how the novice experimenters
approached the uncertainty of measurement by developing a local solution—
the “in-between” strategy—grounded in the requirements of their experiment
and commonsense mathematical reasoning. According to the researchers’
accounts, the procedure made measurement of MT more precise and less arbi-
trary by introducing a systematic way of navigating in a situation of high uncer-
tainty. However, the procedure was also questioned as incorrect by another
laboratory member. The analysis was aimed to provide understanding of the
following: (1) where does local reproducibility of experimental procedures
originate from if it does not originate from rules and (2) how are the actions
identified as correct or incorrect in the shop floor work of practitioners. This
section summarizes how the actions of the research group, which were com-
pletely justifiable in one situation, have later become problematic and discusses
the implications that the analyzed case has for scholarship on reproducibility.

As was demonstrated in the empirical part, the procedure was both
established and dismissed based on how the results of measurement fit in
the main project of practitioners. As in cases of radical uncertainty faced by
experienced ecologists (Roth 2009), the situated work of the experimenters
has been intertwined with the outcomes of the measurement procedure,
which either stabilized previous actions as correct or called them into ques-
tion. In the first three iterations, the results were “good enough” for the
experimenters’ project that required identifying intermediate value and the
spot suitable to continue the experiment. The procedure was dismissed
when it was encountered by the lab member who problematized the adjust-
ments introduced by experimenters to the procedure of MT search and
prompted one of the experimenters to find out that a different search strat-
 egy could result in more precise measurements.

There are strong parallels between identifying mistakes and perceiving
details of the environment as essential or irrelevant (Lindwall, Evans, and
Reit 2017). As has been demonstrated, the invention of the “in-between”
procedure has made specific features of the experimental environment
The intensity of stimulation became noticeable for the experimenters not as just "low" to take a decision but "a bit too low" and thus suitable to execute the "in-between" procedure and take an average threshold. In a similar vein, the requirements of the experiment have highlighted certain details, such as the balance of responses in two muscles, as important, while others, including the number of points checked after identifying a potential threshold, as less essential. To use Michael Lynch’s words, “the very ‘seeing’ of anything at all whether it later is treated as fact or artifact, ‘noise’, or nothing worth bothering about, is itself embedded in a local history of technical activities” (Lynch, Livingston, and Garfinkel 1983, 220). The practice dictates which details, modifications, and variations are counted as significant or unimportant. The contention, as it is depicted in the fourth fragment, appears when participants analyze actions from within two different projects. The reduced time spent on finding the right intensity in a highly uncertain environment was not significant for the research group searching for the intermediate value that would allow them to move to the next experimental step but crucial for a “disengaged” lab assistant who sticks to the meaning of the threshold as “the lowest intensity of stimulation evoking responses in 50 percent of stimulation attempts” and considers an attempt to reduce the number of attempts as being dangerous to the purity of experimental practice.

The criticism of the lab assistant reopened the uncertainty that was closed by the “in-between” method and problematized an attempt to escape the uncertainty with the help of additional criteria specific for the group’s experiment. The deviations between the experimenters’ way of measuring MT and the conventions of laboratory guidelines did not prove a mistake but opened up a discussion around the experimenter’s way of adapting the experimental procedure. The local disagreement was closed only when the outcomes of measurement had confirmed the lab assistant’s criticism and restored certainty by demonstrating that the measurement can be more precise.

The analysis of the connection between uncertainty and disagreement on the laboratory level can be valuable for the discussion on reproducibility across sciences. Science and technology studies have repeatedly stated that reproducibility takes various forms and performs different epistemic functions across disciplines depending on their level of uncertainty (Feest 2019; Leonelli 2018; Peterson and Panofsky 2021). Peterson and Panofsky (2021) show “that fields where routine experimental procedures are less standardized and outcomes are more uncertain tend to pursue more piecemeal and
pragmatic replications, which provides weaker evidence regarding a finding’s veracity” (p. 3). They argue that failed replication efforts have more potential to invalidate the initial results in more standardized fields, even though the scientists there are less likely to initiate these “diagnostic” replications. These are the fields where data sets are shared or the agreement over factors that could influence results exists, such as economics, protein crystallography, or structural biology.

Adding up to the uncertainty argument, I want to highlight the distinction between the embedded uncertainty of scientific work and exposing uncertainty as a strategy for questioning scientific results and methods. Rediscovering uncertainty in the gap between actions and their methodological descriptions is a common strategy used by outsiders of science to question the status of research results (Jordan and Lynch 1998; Lynch 2002). In the analyzed example, it is also a strategy used by the laboratory assistant for reclaiming authority and questioning the “in-between” procedure. The ethnomethodological approach to the problem of rule-following suggests that it is always possible to discover uncertainty on a level of approximation where written rules do not entirely cover practical actions (Garfinkel 1967). Standardization—as a lack of uncertainty—then better serves as an indicator of agreement within the field rather than its inherent characteristic. Viewed in this way, the embedded uncertainty cannot be completely eliminated by courses on statistical analysis or video materials on the craft parts of scientific work, even though they can be valuable means of disseminating knowledge, establishing mainstream approaches and, therefore, stimulating consensus.

What may be problematic in the context of the discussion on questionable research practices is not the embedded uncertainty but the situations when uncertainty is removed from experimental practice with the help of external criteria. In extreme cases, it is possible to imagine how the bias emerges if the criteria stem from other projects such as “publishing an article” or “producing significant results,” which dictate specific requirements for navigating in the terrain of uncertainty. The introduction of new rules will hardly serve as a solution of the problem in the fields with the high level of disagreement, as rules inevitably require more rules and, and, as discovered in the study of “questionable research practices,” are still not enough to distinguish between potentially dangerous violations and normal research conduct (Fiedler and Schwarz 2016). Changing the system of stimulus, which overstimulates the production of publishable results, seems to be a more efficient solution to the problem than external control of replicability. Identifying the logic of practice, which makes certain
Procedures a reasonable way of overcoming uncertainty, can be an alternative enterprise, shifting the focus from the inevitable gap between actions and formal descriptions to the question of whether the researchers following the one set of instructions are engaged in the same project.

The second central issue of this paper was the question of where does the reproducibility of practices originate from, if not guidelines that are inevitably incomplete and require engaging additional “ad hoc” considerations (Garfinkel 1967) and background knowledge to guide practitioners through their blind spots (Amerine and Bilmes 1988)? The analysis of four chronologically recorded fragments aimed to demonstrate how the local reproducibility of the “in-between” measurement procedure is produced in situ, that is, how the experimenters make recognizable for themselves and for us as observers that they are engaged in the new iterations of the same procedure. It was demonstrated that the participants establish continuity through the reenactment of the “in-between” procedure’s constitutive features—“substrates” (Goodwin 2018)—that are reproduced throughout all the iterations when the procedure is used. The most obvious substrate is the name of the procedure (“in-between”) appearing in the first, second, and fourth fragment and making the similarity between the iterations plain to see. However, the name is not the only constant feature of the excerpts; the others are structural elements of the procedure, most clearly explicated during its second use: the definition of the situation, which presupposes the use of the “in-between” procedure; the threshold value as its result; and implication as the next procedural step of calculating 120 percent. The reproduction of these substrates allows both analysts and researchers identify “same but different” iterations of the habitual procedure, ensuring its recognizability as an established approach. The references become subtle with the establishment of the procedure but remained visible though the patterns of movements, the order of speech, and the sequential organization of situation. These patterns make the application of the “in-between” procedure recognizable and ensure its local reproducibility.

As reproducibility is an inevitably local achievement, the formal answer to the question of its production is quite limited. Reproducibility is produced in situ through the subtle references, imitation, and copying. It is achieved in the “seen but unnoticeable” (Garfinkel 1967, 180) details of local order production, which do not and cannot appear on the aggregated level of formal methodological descriptions of scientific practice. At the same time, neglecting the local origins of reproducibility can lead to conceptual confusions and reinforce the gap between actual work of scientists and its theoretical display.
Appendix

Transcription Conventions

(0.3) – the length of a pause in the tenths of seconds
(.) – short pause, around the tenth of a second
() – action
(() – transcriber’s comments
[] – overlaps
: – lengthy sound
* * – the utterance pronounced in a lower voice
= absence of pause between turns
__ – stressed utterance
<> – the utterance produced slowly
>< – the utterance produced faster

Acknowledgments

The paper was written during my research visit to the University of Gothenburg funded by the Swedish Institute. I am grateful to Oskar Lindwall for his generous help in developing this paper. I thank Mick Smith, Andrei Korbut, participants of the SDS seminar at the University of Gothenburg, editors, and the anonymous reviewers for their useful suggestions on improving this article. Last but not least, I am thankful to the laboratory members who allowed me to follow their work and found the time to explain the details of their practice.

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 the following financial support for the research, authorship, and/or publication of this article: This article received funding from Swedish Institute

ORCID iD

Kristina Popova https://orcid.org/0000-0002-4320-4578

Notes

1. “As an example, consider a very simple decision faced by researchers analyzing reaction times: how to treat outliers. […] Most (but not all) researchers excluded some responses for being too fast, but what constituted “too fast” varied
enormously: the fastest 2.5%, or faster than 2 standard deviations from the mean, or faster than 100 or 150 or 200 or 300 ms. Similarly, what constituted “too slow” varied enormously: the slowest 2.5% or 10%, or 2 or 2.5 or 3 standard deviations slower than the mean, or 1.5 standard deviations slower from that condition’s mean, or slower than 1,000 or 1,200 or 1,500 or 2,000 or 3,000 or 5,000 ms. None of these decisions is necessarily incorrect, but that fact makes any of them justifiable and hence potential fodder for self-serving justifications” (Simmons, Nelson, and Simonsohn 2011, 1360).

2. The participants’ names are changed.

References
Alac, M. 2011. Handling Digital Brains: A Laboratory Study of Multimodal Semiotic Interaction in the Age of Computers. Cambridge, MA: The MIT Press.

Amerine, R., and J. Bilmes. 1988. “Following Instructions.” Human Studies 339 (11): 327-39.

Bjelic, D., and M. Lynch. 1992. “The Work of a (Scientific)Demonstration: Respecifying Newton’s and Goeth’s Theories of Prismatic Color.” In Text in Context: Contributions to Ethnomethodology, edited by G. Watson and R. M. Seiler, 52-78. London, UK: Sage.

Button, K. S., J. P. A. Ioannidis, C. Mokrysz, B. A. Nosek, J. Flint, E. S. J. Robinson, and M. R. Munafò. 2013. “Power Failure: Why Small Sample Size Undermines the Reliability of Neuroscience.” Nature Reviews Neuroscience 14 (5): 365-76.

Collins, H. M. 1989. Changing Order: Replication and Induction in Scientific Practice. London, UK: Sage.

Collins, H. M. 2001. “Tacit Knowledge, Trust and the Q of Sapphire.” Social Studies of Science 31 (1): 71-85.

Darling, W. G., and A. J. Butler. 2006. “Stimulation Depends on Muscle Activation.” Experimental Brain Research 174 (2): 376-85.

Delamont, S., and P. Atkinson. 2001. “Doctoral Uncertainty: Mastering Craft Knowledge.” Social Studies of Science 31 (1): 87-107.

Dreyfus, H., and S. Dreyfus. 1980. “A Five-stage Model of Mental Activities Involved in Directed Skill Acquisition.” Supported by the U.S. Air Force, Office of Scientific Research (AFSC) under contract F49620-C-0063 with the University of California, Berkeley, CA.

Feest, U. 2019. “Why Replication Is Overrated.” Philosophy of Science 86 (5): 895-905.

Fiedler, K., and N. Schwarz. 2016. “Questionable Research Practices Revisited.” Social Psychological and Personality Science 7 (1): 45-52.
Freese, J., and D. Peterson. 2018. “The Emergence of Statistical Objectivity: Changing Ideas of Epistemic Vice and Virtue in Science.” Sociological Theory 36 (3): 289-313.

Garfinkel, H. 1967. Studies in Ethnomethodology. Englewood Cliffs, NJ: Prentice Hall.

Garfinkel, H. 2002. Ethnomethodology’s Program: Working Out Durkheim’s Aphorism. Edited by A. Rawls. Lanham, MD: Rowman and Littlefield.

Garfinkel, H., and K. Liberman. 2007. “Lebenswelt Origins of the Sciences.” Human Studies 30 (3-7): 3-56.

Garfinkel, H., M. Lynch, and E. Livingston. 1981. “The Work of a Discovering Science Constructed with Materials from the Optically Discovered Pulsar.” Philosophy of the Social Sciences 11 (2): 131-58.

Goodman, S. N., D. Fanelli, and J. P. A. Ioannidis. 2016. “What Does Research Reproducibility Mean?” Science Translational Medicine 8 (341): 341ps12.

Goodwin, C. 1994. “Professional Vision.” American Anthropologist 96 (3): 606-33.

Goodwin, C. 2000. “Practices of Color Classification.” Mind, Culture, and Activity 7 (1–2): 19-36.

Goodwin, C. 2018. Co-operative Action. New York: Cambridge University Press.

Hallett, M. 2000. “Transcranial Magnetic Stimulation and the Human Brain.” Nature 406 (6792): 147-50.

Heath, C., J. Hindmarsh, and P. Luff. 2010. Video in Qualitative Research. London, UK: Sage.

Herbsman, T.L. Forster, C. Molnar, R. Dougherty, D. Christie, D. Ramsey, J. Koola, et al. 2009. “Motor Threshold in Transcranial Magnetic Stimulation: The Impact of White Matter Fiber Orientation and Skull-to-cortex Distance.” Human Brain Mapping 30 (7): 2044-55.

John, L. K., G. Loewenstein, and D. Prelec. 2012. “Measuring the Prevalence of Questionable Research Practices with Incentives for Truth Telling.” Psychological Science 23 (5): 524-32.

Jordan, K., and M. Lynch. 1998. “The Dissemination, Standardization and Routinization of a Molecular Biological Technique.” Social Studies of Science 28 (5-6): 773-800.

Knorr-Cetina, K. 1981. The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science. Oxford, UK: Pergamon Press.

Latour, B., and S. Woolgar. 1979. Laboratory Life. The Construction of Scientific Facts. Princeton, NJ: Princeton University Press.

Leahey, E. 2008. “The Case of Data Editing.” Science, Technology, & Human Values 33 (5): 605-30.

Leonelli, S. 2018. “Rethinking Reproducibility as a Criterion for Research Quality.” In Curiosity, Imagination, and Surprise, edited by L. Fiorito, S. Scheal, and C. E. Suprinyak, 129-46. Bingley, UK: Emerald.
Lindwall, O., B. Evans, and C. Reit. 2017. “The Recognizability and Consequenti-
ality of Mistakes: Some Notes on the Article by Klemp et al.” Éducation & Didactique 11 (1): 133-38.
Lynch, M. 1985. Art and Artifact in Laboratory Science: A Study of Shop Work and
Shop Talk in A Research Laboratory. London, UK: Routledge & Kegan Paul.
Lynch, M. 1993. Scientific Practice and Ordinary Action. New York: Cambridge
University Press.
Lynch, M. 2002. “Protocols, Practices, and the Reproduction of Technique in
Molecular Biology.” The British Journal of Sociology 53 (2): 203-20.
Lynch, M., E. Livingston, and H. Garfinkel. 1983. “Temporal Order in Laboratory
Work.” In Science Observed: Perspectives on the Social Study of Science, edited
by K. Knorr-Cetina and M. Mulkay, 205-38. London, UK: Sage.
Ogien, A. 2009. “Rules and Details.” Journal of Classical Sociology 9 (4): 449-73.
Open Science Collaboration. 2015. “Estimating the Reproducibility of Psychologi-
cal Science.” Science 349 (6251): aac4716.
Peterson, D. 2016. “The Baby Factory.” Socius: Sociological Research for a
Dynamic World 1 (2): 1-10.
Peterson, D., and A. Panofsky. 2021. “Self-correction in Science: The Diagnostic
and Integrative Moves for Replication.” Social Studies of Science. doi: 10.1177/
03063127211005551
Polanyi, M. 1983. The Tacit Dimension. Gloucester, MA: Peter Smith.
Rossi, S., M. Hallett, P. M. Rossini, and A. Pascual-Leone, Safety of TMS
Consensus Group. 2009. “Safety, Ethical Considerations, and Application
Guidelines for the Use of Transcranial Magnetic Stimulation in Clinical
Practice and Research.” Clinical Neurophysiology 120 (12): 2008-39.
Rossini, P. M.A. T. Barker, A. Berardelli, M. D. Caramia, G. Caruso, R. Q. Cracco,
M. R. Dimitrijevic, et al. 1994. “Non-invasive Electrical and Magnetic Stimula-
tion of the Brain, Spinal Cord and Roots: Basic Principles and Procedures for
Routine Clinical Application.” Electroencephalography and Clinical Neurophy-
siology 91 (2): 79-92.
Roth, W.-M. 2009. “Radical Uncertainty in Scientific Discovery Work.” Science,
Technology, & Human Values 34 (3): 313-36.
Roth, W.-M., and G. M. Bowen. 2001. “‘Creative Solutions’ and ‘Fibbing Results’:
Enculturation in Field Ecology.” Social Studies of Science 31 (4): 533-56.
Shutter, D., and J. Honk. 2006. “A Standardized Motor Threshold Estimation
Procedure for Transcranial Magnetic Stimulation Research.” Journal of ECT
22 (3): 176-78.
Simmons, J. P., L. D. Nelson, and U. Simonsohn. 2011. “False-positive Psychology:
Undisclosed Flexibility in Data Collection and Analysis Allows Presenting
Anything as Significant.” Psychological Science 22 (11): 1359-66.
Sommer, M., T. Wu, F. Tergau, and W. Paulus. 2002. “Intra- and Interindividual Variability of Motor Responses to Repetitive Transcranial Magnetic Stimulation.” *Clinical Neurophysiology* 113 (2): 265-69.

Sormani, P. 2014. *Respecifying Lab Ethnography: An Ethnomethodological Study of Experimental Physics.* Surrey, UK: Ashgate.

Stevens, J. R. 2017. “Replicability and Reproducibility in Comparative Psychology.” *Frontiers in Psychology* 8:1-6.

Suchman, L. A. 1987. *Plans and Situated Actions: The Problem of Human-machine Communication.* Cambridge, UK: Cambridge University Press.

Tranulis, C., B. Guéguen, A. Pham-Scottez, and M. N. Vacheron. 2006. “Motor Threshold in Transcranial Magnetic Stimulation: Comparison of Three Estimation Methods.” *Neurophysiologie Clinique [Clinical Neurophysiology]* 36 (1): 1-7.

**Author Biography**

Kristina Popova is a PhD candidate at KTH Royal Institute of Technology in Stockholm. This paper was written during her fellowship in EMCA_RU research group based in Moscow where she studied rule-following at workplaces with a focus on laboratory work in cognitive neuroscience.