View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Crossref

Math Ed Res J (2017) 29:455–470
DOI 10.1007/s13394-016-0187-x

ORIGINAL ARTICLE

An inferentialist perspective on the coordination of actions and reasons involved in making a statistical inference

Arthur Bakker[1] · **Dani Ben-Zvi**[2] · **Katie Makar**[3]

Received: 24 June 2016 / Revised: 28 December 2016 / Accepted: 30 December 2016 /
Published online: 31 January 2017
© The Author(s) 2017. This article is published with open access at Springerlink.com

Abstract To understand how statistical and other types of reasoning are coordinated with actions to reduce uncertainty, we conducted a case study in vocational education that involved statistical hypothesis testing. We analyzed an intern's research project in a hospital laboratory in which reducing uncertainties was crucial to make a valid statistical inference. In his project, the intern, Sam, investigated whether patients' blood could be sent through pneumatic post without influencing the measurement of particular blood components. We asked, in the process of making a statistical inference, how are reasons and actions coordinated to reduce uncertainty? For the analysis, we used the semantic theory of inferentialism, specifically, the concept of webs of reasons and actions—complexes of interconnected reasons for facts and actions; these reasons include premises and conclusions, inferential relations, implications, motives for action, and utility of tools for specific purposes in a particular context. Analysis of interviews with Sam, his supervisor and teacher as well as video data of Sam in the classroom showed that many of Sam's actions aimed to reduce variability, rule out errors, and thus reduce uncertainties so as to arrive at a valid inference. Interestingly, the decisive factor was not the outcome of a *t* test but of the reference change value, a clinical chemical measure of analytic and biological variability. With insights from this case study, we expect that students can be better supported in connecting statistics with context and in dealing with uncertainty.

✉ Arthur Bakker
a.bakker4@uu.nl

[1] Freudenthal Institute, Utrecht University, Utrecht, The Netherlands

[2] Faculty of Education, The University of Haifa, Haifa, Israel

[3] School of Education, The University of Queensland, St Lucia, Australia

Keywords Inferentialism · Laboratory education · Statistical inference · Uncertainty · Vocational education · Webs of reasons and actions

The endless cycle of idea and action,

Endless invention, endless experiment

T. S. Eliot

The aim of this article is to show how different types of reasoning (including statistical reasoning) and actions are coordinated in making a statistical inference in a real-life context with a focus on reducing uncertainty. This requires a theoretical framework that convincingly shows how knowledge (including the ability to reason) and action are coordinated. To this end, we deploy inferentialism—the topic of the special issue.

In statistics and mathematics education, several researchers have already explored the value of inferentialism (e.g., Bakker and Derry 2011; Hußmann and Schacht 2009; Pratt 2012; Schindler and Hußmann 2013). Bakker and Akkerman (2014) propose that inferentialism may provide a useful theoretical framework to analyze vocational and workplace knowledge as coordinating various types of reasoning and actions (see also Guile 2006). In the current article, we suggest that the different types of reasoning (practical, mathematical, statistical, and chemical) and actions become coordinated through inferential relations into webs of reasons and actions. Such reasons include premises and conclusions, implications, motives for action, and utility of tools for particular purposes in a particular context (cf. Bakker and Derry 2011). We think that an analysis of webs of reasons and actions involved in a particular case can provide insight into how types of reasoning and actions are coordinated and can eventually help to improve the teaching of statistical reasoning in real-life contexts.

It is still not clear enough how students learn to make statistical inferences in context (Makar et al. 2011). In line with this practical concern, more theoretical and empirical study is needed on the relations between statistical inference and its application in practice. To this end, we focus on the statistical inferences made in a vocational setting. The study of such settings helps better understand what students need to learn and how learning environments focusing on statistical inference can be improved. We present a case study in an area of education that has so far received little attention, vocational education. Rather than a narrow focus on statistical reasoning or inference, this area requires a comprehensive approach in which attention to statistical or probabilistic reasoning should be combined with attention to other disciplinary reasoning (e.g., chemical) as well as actions. As such, the case presented is one of the reasoning and actions required to reduce uncertainty so as to arrive at a valid statistical inference. Insights generated from the case study are also relevant for general education that takes context or laboratory work seriously (e.g., Ben-Zvi and Aridor 2016; Dierdorp et al. 2011; Heinicke and Heering 2013; Makar et al. 2011).

Theoretical background

Statistical inference

Statistical inference is at the core of statistics because it allows people to draw conclusions from data about a wider universe such as a population or process (Moore 1995, p. 3). Although students have difficulty with formal statistical inference, many aspects of inference are relevant for everyday use—making predictions, estimating based on available information, and judging the reasonableness of a solution or a claim all involve making inferences—as well as in workplaces (Bakker et al. 2008). It has also been acknowledged that informal statistical inference is to be used with nonspecialists such as school children, introductory statistics students, and in workplace settings (Wild et al. 2011).

Over the last decade, the statistics education research community has made considerable progress on the theme of statistical inference (Makar and Rubin 2009) and the informal inferential reasoning that underlies statistical inference (Ben-Zvi 2006; Makar et al. 2011; Pfannkuch 2006; Zieffler et al. 2008). Makar and Rubin (2009) identified three features that characterize a statistical inference: (1) a statement of generalization beyond the data, (2) use of data as evidence to support this generalization, and (3) the use of probabilistic (non-deterministic) language that expresses some uncertainty about the generalization. These features apply to both formal techniques such as hypothesis testing (e.g., using a t test) or point estimation and students' beginning informal conceptions of statistical inference (Wild et al. 2011).

Thus one key feature of statistical inference is the uncertainty involved. Statistics education typically deals with uncertainty in terms of confidence intervals, probabilities or p values. These statistical concepts are known to be difficult for tertiary students (Garfield and Ben-Zvi 2008). What is known much less about is how to deal with the uncertainties involved in making an inference in the broader context of a modeling task, risk assessment, or an authentic project. In such cases, the types of uncertainty that are not necessarily quantifiable need to be taken into account, such as the quality of the design or of the measurement, and hence, the quality of the data from which the statistical inference can be made (cf. Arnold et al. 2013). Our purpose of mentioning these types is to emphasize that in our case study, we intend to stay open to types of uncertainty that seemed relevant in the actions taken to arrive at a valid inference.

Webs of reasons and actions

Several psychologists (e.g., Piaget 1970/2013) and philosophers have addressed the relationship between knowledge and action. Dewey (1929/2008), for example, is known for his analysis of the human quest for certainty and his explanation for why education often treats knowledge as more important than action. This hierarchy is still prominent today. For example, in learning to experiment in physics and chemistry education, university students' laboratory skills are generally poor. This is presumably due to their education focusing purely on theoretical knowledge (Heinicke and Heering 2013). In vocational education, it is common for students to complain on the knowledge taught in courses as being too abstract (Roth 2014; Wedege 1999).

As announced, we focus on the work of a more recent philosopher whose work we consider especially relevant to vocational education, but also general education: Robert Brandom. Like Dewey, Brandom treats knowledge and action democratically (see also Brandom 2008), without any prior assumptions about hierarchy. This is evident in his discussion of action and judgment (a general term for statements, propositions, claims, conclusions, and inferences). In line with Kant, Brandom (2000, p. 159) considers judgments to be the minimal unit that one can be responsible for at a cognitive level, just like actions are the minimal unit one can be responsible for at a practical level. Following American pragmatists such as Dewey, Brandom has a pragmatist view on concepts:

> To grasp or understand (…) a concept is to have practical mastery over the inferences it is involved in—to know, in the practical sense of being able to distinguish, what follows from the applicability of a concept, and what it follows from. (Brandom 2000, p. 48)

We should note that Brandom's technical usage of the word inference is much broader than the typical philosophical meaning (relation between a premise and a conclusion) and the statistical meaning (a conclusion about some wider universe). For Brandom, almost any judgment, even the smallest, is an inference. For example, "fire!" is already part of an inferential network including the risk of a life-threatening situation, which explains why a human would express such a statement (for a philosophical discussion of noninferential statements, see Brandom 2002).

Thus, Brandom, in Hegelian spirit, puts emphasis on the inferential nature of knowledge: "Claims both serve as and stand in need of reasons or justifications. They have the contents they have in part in virtue of the role they play in a network of inferences" (Brandom 2000, p. 162). This inferential nature is privileged over the representational nature of knowledge—an idea that has been eloquently formulated by one of Brandom's inspirators, Wilfrid Sellars (1956, §36):

> in characterizing an episode or a state as that of *knowing*, we are not giving an empirical description of that episode or state; we are placing it in the logical space of reasons, of justifying and being able to justify what one says.

Philosophers use the term space of reasons for anything conceptual that should be distinguished from the realm of law about causes and effects (McDowell 1996). For our study of the reasons for actions that people perform as part of their work, the term web of reasons (Brandom 1994, p. 5) seems more useful. A web of reasons can be characterized as a complex of interconnected reasons; these reasons include premises and conclusions, inferential relations, implications, motives for action, and utility of tools for particular purposes in a particular context (Bakker and Derry 2011). In the workplace settings, reasons can be of a different nature: some are practical and some are theoretical, often weighted by their relative merits. Bakker et al. (2008) give an example from the automotive production industry in which practical reasons outweighed theoretical reasons. Although not statistically sound, the employees' decisions made sense in the light of the situation, ultimately avoiding any

customer complaints. Bakker et al.'s analysis suggests that vocational students, apprentices, and employees need to learn to reason with a web of multiple relevant reasons (practical, statistical, and mathematical) when making action-oriented decisions based on the projected implications for their practical contexts.

Based on inferentialist ideas, we propose that coordination between knowledge (as actualized or expressed in facts, statements, judgments, and conclusions) and action can be constituted by reasons. Therefore, we use the term webs of reasons and actions. Reasons are relational: R is a reason for S, where R and S can refer not only to judgments (facts and claims) but also actions and even feelings (something can be done to become happy). In line with the philosophical literature, we use the term reason in a broad sense (McDowell 1996): a premise has a reasonable or inferential relation with a conclusion and a motive with an action. In cognitive linguistics (cf. Sanders et al. 1992), many of such inferential relations are distinguished, for example:

- P is a premise for conclusion C,
- W is a warrant for statement S,
- M is a means to the end E,
- X is a motive for action A,
- T is a tool to get A done,
- D is done for the purpose of Q.

Note that many of the things related are judgments (P, S, and W), but some can be actions or even feelings (A, Q, and X). In cognitive linguistics, these inferential relations are considered interesting because they make texts coherent. We argue more generally that such reasonable or inferential relations help to coordinate different types of knowledge or reasoning (statistical mathematical, practical, and chemical). As such, reasons and inferential relations seem to form the "coordination glue."

With this theoretical background in mind, we expect to see in a vocational student's research project how diverse knowledge types and actions are inferentially related. Having explained the main concepts in our case study (statistical inference and webs of reasons and actions), we can now formulate our research question: in the process of making a statistical inference, how are reasons and actions coordinated to reduce uncertainty?

Methods

Participant and setting

The research presented here is a case study of Sam's research project as part of his internship in a hospital laboratory. The case is one of the knowledge and actions required to reduce uncertainty so as to arrive at a valid statistical inference (not Sam's learning process). Sam (pseudonym) is 19 years old and attends the highest level (4) of a Dutch senior secondary vocational laboratory school (MBO) with which Bakker had prolonged engagement through both survey and design-based research (Bakker 2014b; Bakker and Akkerman

2014). This level of vocational education is below bachelor level; but with a diploma, Sam would be entitled to enter higher professional education at a bachelor level. This type of vocational education starts with primarily full-time school-based education and ends with primarily full-time internships. During the last year, students come back to school for 1 day every 2 weeks. It was during such release days that the first author stimulated students to refresh their statistics and link it to their research projects (see Bakker and Akkerman 2014 for the first design-based cycle in this school). The students had mainly learned basic statistical concepts and procedures in the first 2 years with more complex content in their third year such as correlation, regression, coefficient of variation, and *t* test, but also more dedicated techniques such as statistical process control. The *t* test is typically addressed in just one lesson (see Bakker 2014b). According to the teachers and supervisors, most students of laboratory education neither recall what they learned about statistics nor know how to apply it.

This design-based research setting provided us with the opportunity to do a more in-depth case study of Sam's research project; in this case, with a focus on uncertainty. For this purpose, we deliberately focused on the webs of reasons and actions about uncertainties involved in his project and the actions to reduce these uncertainties, which eventually led to a statistical inference. We have chosen Sam for this case study because his project is more easily conveyed to medical nonspecialists than those of the other students.

Sam considered himself a "lousy student." He often came late, did not have books or notes with him, and was not motivated to become a laboratory technician.

Sam's research project

Like companies, hospitals have to be efficient with their resources and, therefore, use modern process improvement techniques such as Six Sigma or Lean production to reduce unnecessary use of time and money (Womack et al. 1991). One of the questions that Sam's work supervisor had was whether patient blood could be sent through the pneumatic post more often rather than be brought by foot to the hospital laboratory for blood testing. Hospitals use pneumatic post, tube systems, through which they send materials quickly to their destinations within the hospital. Sending patients' blood samples to the laboratory by pneumatic post saves time. This is done for blood component measurements that are considered to be relatively unaffected by the post, for example, LDH (lactate dehydrogenase). However, some blood components are known to be potentially affected by shaking or shocks. Shaking can lead to hemolysis, the process of damaging the red blood cells (erythrocytes) and releasing their content into the blood plasma. In red blood cells, the concentration of potassium (Na^+), for example, is high (about 140 mmol/L), but low in the plasma (about 4 mmol/L). Potassium is typically measured in the plasma, so if hemolysis occurs, the measurement of potassium concentration is escalated. Similar problems may occur for other components such as thrombocytes. Shocks can also initiate coagulation which would also lead to misleading measurements. This is why these components were being measured after bringing the samples on foot to the laboratory through the hospital corridors rather than relying on the faster pneumatic post.

One may expect that researchers had already studied the effects of pneumatic posts in hospitals. However, we were told that pneumatic posts can be quite different and a few general rules hold. Sam's task was to check whether the concentrations of several blood components could be affected by the pneumatic post in the hospital where he was doing his internship. We emphasize that this was a genuine research project in the sense that the hospital did not know the effects of the pneumatic post on various blood components and that Sam's results had a real impact on the hospital's policy.

In designing his study, which served as an assessment in a course within his vocational education program, Sam had to devise his own initial plan, but he was guided by laboratory technicians and a clinical chemist in working out the details of his research project.

Data collection

The main data sources are video recordings of six 1-h meetings that the first author had with three students and their teacher in the vocational laboratory school and video recordings of Sam's final presentation (30 min) and post-interview (20 min). We also have audio recorded interview data with Sam and his workplace supervisor. In the interviews, we made sure to ask for reasons for actions. All recordings were transcribed verbatim. The data were collected by the first author, but analyzed collaboratively by all three authors. Translations into English were done by a professional translator and checked by the first author. We have both Sam and his supervisor's permission to use the data as collected by Sam.

Data analysis

Our analysis aimed to identify the reasons and actions involved in Sam's research project to reduce uncertainty and arrive at a valid statistical inference. This required several steps. The first was to become familiar with the relevant disciplinary knowledge required to understand the project (chemistry and practicalities of laboratory work). The second step was to identify all actions and possible reasons for them as mentioned in the transcripts and selected relevant parts of the meetings, presentation, and interviews. For example, when Sam stated that citrate tubes had to be used first (action A3.1 in Fig. 1) when tapping blood, he explained the reason (summarized as R3.1) as follows:

> There's a small rubber, you have a needle, then you have a cover in which sits a small pricking device, which pierces through the rubber of the tubes. Sometimes there's already anti-coagulation stuff at the rubber of this tube and that is being taken along by that pricking device to the next tube.

Conjectures and conclusions about the actions and reasons were jointly discussed until agreement was reached. In cases of doubt we verified conclusions with Sam or his teacher, or with data collected at a later stage. Last we checked our analysis with Sam's workplace supervisor. In this way, as a last step, we were able to summarize webs of reasons and actions related to uncertainties and inferences involved in his project which provide an answer to the research question of how reasons and actions were coordinated.

Fig. 1 A web of reasons and actions involved in Sam's experiment. *A* stands for action and *R* for reason. The action-reason combinations are presented clockwise in chronological order. The more detailed action-reason combinations are presented more toward the periphery

Case study

One statistical inference

We start the case study with a key statistical inference of Sam's report. Although it was common practice to send LDH samples by pneumatic post, it turned out that the LDH values measured in the pneumatic post condition were somewhat higher on average than those in the walking condition. Comparison by means of a t test yielded a p value of .0017. When focusing on the statistical inference, one may conclude that because the difference between the two conditions was statistically significant, hemolysis has occurred. It would therefore not be wise to send the blood for LDH measurements through the pneumatic post. However, the case study will show how Sam incorporated many more considerations than the p value in his final conclusion and how his conclusion was different from what we had expected based on the statistical inference. A more holistic approach is required to

understand how Sam carried out an experiment (action A in Fig. 1) to arrive at valid inference (reason R). In order to make a valid inference, Sam needed to collect valid data. We summarize several actions and their underlying reasons to illustrate how he reduced uncertainty due to the quality of the design and the data. Some of these we represent as webs to illustrate our point of the web-like nature of Sam's actions and reasons (Fig. 1). We have put the main action in the middle. The actions are ordered clockwise in chronological order from A1 to A8. In activity theory, it is common to distinguish between activity (e.g., the overall experiment), action, and operation, but for our purpose of emphasizing the reasons at stake, we do not need this distinction here.

Collecting valid data

The collection of valid data involved several actions. First, a sample of people had to be selected (action A1) given the need for data to make an inference (reason R1a) and impossibility to study a population of patients (reason R1b). Therefore, Sam had to decide on the sample size (action A1.1). Considerations (reasons) for determining sample size were: large enough to make a valid inference (reason R1.1a) but small enough to minimize the number of people who would be burdened with giving blood (reason R1.1b). Sam selected healthy colleagues (action A1.2) so as not to unnecessarily burden patients with drawing blood (reason R1.2a) and because colleagues are easily accessible (R1.2b). Moreover, the issue to be studied apparently did not depend on people's health (a reason not represented in the figure).

A blood test is a laboratory analysis performed on a blood sample to determine physiological and biochemical states, such as disease or mineral content. A blood sample is usually drawn from a vein in the arm using a needle. Sam drew the blood samples himself (action A2) to exclude some possible sources of variability (reason R2). For example, he makes the band "not too tight" (action A2.1) to ensure (reason R2.1) that particular processes (e.g., hemolysis) do not occur that would distort the concentrations to be measured. If the elastic band wrapped around the upper arm were too tight, blood may be injected too fast so that blood cells would get damaged. Concentrations of particular substances can seem to be too high because particles have left the blood cells (hemolysis). Sam was aware of the proper blood sampling procedures and underlying reasons.

There are further rules for taking blood samples for particular measurements in a particular order for each patient (action A3) because some substances may be contaminated by anticoagulants in the rubber ring of particular tubes (reason R3). As Sam explained after his final presentation (see quote in the "Data analysis" section), the citrate tubes, without anticoagulants, have to be used first. The blood tubes rings that do contain anticoagulants are used after the citrate tubes (action A3.1); otherwise, the presence of these substances could influence the measurement of other components (reason R3.1).

Sam only collected blood samples of five colleagues each day (action A4) because more samples could not be analyzed on the same day (reason R4a). Moreover, storing blood in a freezer or waiting too long could lead to changing concentrations and distorted measurements (reason R4b) and, thus, invalid inferences (underlying reason R).

The two conditions

In line with the principles of a split sample experiment, each blood sample was split into two equal halves. Half of the samples were brought by foot and half were sent by the pneumatic post (action A5) to ensure that differences in measurements could only be attributed to difference in condition (reason R5). From a clinical view, the order in which the two samples of tubes are sent does not matter; but to rule out any potential order effects (reason R5.1), Sam alternately brought the first by foot or sent it by post (action A5.1). For the experimental condition, Sam used the longest tube in the pneumatic post (action A6) to have the most extreme case (speeds up to 8 m/s and *g* values above 10); this would allow him to infer that if no difference were found, sending blood by post would be safe because other postings would be under less extreme conditions (reason R6).

Measurement and calibration

For each patient, Sam analyzed the samples in the same order (action A7). In this way, Sam reduced any variability potentially arising from reordering (reason R7). The blood measurement machine is calibrated (action A8) on a daily basis because otherwise the machine may produce wrong values (R8): measurement machines typically start to become less accurate after some time due to small mechanical changes (cf. Bakker et al. 2011). However, Sam did not have to do this himself; it was done by the regular laboratory technicians (A8.1) because this happened as part of the daily routine of the laboratory anyway (reason R8.1).

Having carried out the experiment, Sam had made the aforementioned statistical inference using a t test yielding a p value of .0017. For those trained in hypothesis testing, this result may be reason for a simple action: the decision to keep walking the blood to the laboratory rather than sending it by pneumatic post. After all the significant difference points to hemolysis taking place otherwise. However, statistical significance is only informative with an indication of the accompanying effect size (Ellis 2010). If the effect size has not practical significance, then the statistical significance is of little importance to decision-making. Sam did not use an effect size but a reference change value (RCV), a clinical chemical measure of analytic and biological variability used in laboratory settings that was also produced on the data generated in the experiment. This RCV stayed within the limits of what counts as unproblematic variation in laboratory and, hence, the decision made by the hospital was to send blood by pneumatic post even if Potassium was to be measured. As a consequence, much time was saved, and nurses who beforehand had to walk blood to the laboratory were happy. In Sam's own words:

> They [the hospital administrators] were already very happy with these measurements because they are now not allowed to go via tube mail, and it is allowed what comes out of here. Then the departments are very happy, then they do not have to run each time.

In Fig. 2, we have summarized how webs of reason and actions can be expanded so as to include such emotions (which in turn can be reasons for actions). Of course, many

more emotions played a role: for example, it was clear from the interviews that Sam was proud of his project.

Discussion

As part of a larger research agenda of trying to understand how actions and different types of reasoning (statistical, mathematical, chemical, clinical, and practical) can be coordinated, especially in vocational educational settings, we carried out a case study of the knowledge and actions required to reduce uncertainty so as to arrive at a valid statistical inference. The case presented as a research project carried out by an intern in a hospital laboratory. Using an inferentialist perspective and language, we suggest that reasons are inferential relations that can provide the coherence between actions and types of knowledge, here with a specific interest in statistics in context. We asked, in the process of making a statistical inference, how are reasons and actions coordinated to reduce uncertainty?

To highlight the nature of such coordination, our analysis focused on webs of reasons and actions. Some laboratory procedures are well defined and prescribed, and reasons for them were rather obvious (sampling of blood). Some actions were based on a wealth of clinical and chemical knowledge oriented toward valid statistical inference.

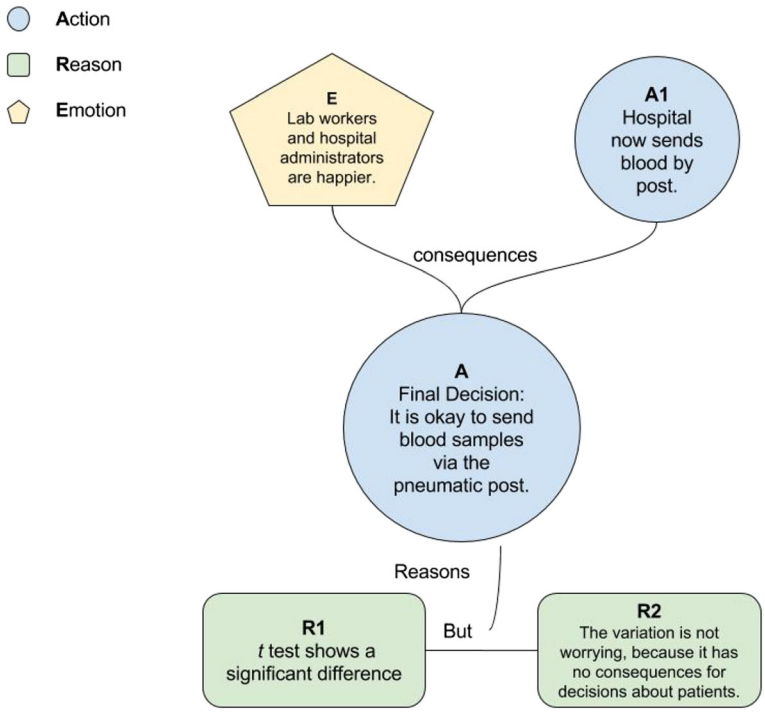

Fig. 2 Web of reasons and actions regarding the final decision

Although we have focused on actions and reasons, there were also emotions involved. Sam seemed proud about his surprising results, and the hospital supervisor and nurses were happy with the results as they saved employees walking and the hospital money (Fig. 2).

The case study points to several more specific conclusions. First, the hypothesis testing (in this case, *t* tests) typically taught in introductory statistics courses only played a subsidiary role in Sam's analysis. In the end, the clinical relevance of the chemical measure of RCV was a more decisive factor than the *t* test outcomes in deciding whether blood samples could be sent through the pneumatic post. Statistics courses teach what a *t* test is for and under what assumptions it can be used, but may be limited in helping students to understand under what conditions such a test is helpful and what conclusions can be drawn. The inferentialist and pragmatist view on concepts cited earlier implies that statistics educators should help students "know, in the practical sense of being able to distinguish, what follows from the applicability of a concept, and what it follows from" (Brandom 2000, p. 48). We, as authors, initially without the clinical-chemical background required, were inclined to conclude that LDH measurement should not be carried out on blood samples sent through the pneumatic post. If statistics educators want to point out that statistical significance may not be practically relevant, they may refer to large sample size or the importance of effect size. However, even differences with a medium effect size may be practically not very relevant. Sam's supervisor told us that the amount of hemolysis pointed to by the statistically significant outcome is so small that it would not lead to clinically wrong advice to medical specialists. He based this view on the interpretation of the RCV. In our view, this highlights the need for a holistic view on concepts and techniques, and shows that integration of different types of knowledge is required to draw sensible conclusions (cf. Bakker and Akkerman 2014).

The second point we want to make is that the actions and underlying reasons mentioned in our case study are of course only some of the more prominent ones. The web of reasons and actions we included here (Fig. 1) was but one portion of Sam's project. It seems impossible to represent full webs of reasons and actions involved: for every fact, judgment, emotion, and action, there are underlying reasons, and participants in the practice under investigation may even not be aware of many of them. We have confined our analysis to those explicitly mentioned by Sam and validated with his supervisor. In our view, Sam's awareness of the main reasons was rather well coordinated, although we understood his project better after having interviewed his supervisor. Despite Sam's self-image of a "lousy student," he had quite detailed knowledge about the practical, chemical, and statistical issues involved in relation to what was clinically relevant. He used techniques that are well beyond anything taught in general secondary education (*t* tests are not taught in the Dutch mathematics curriculum and neither is RCV part of the science curriculum). As such, his knowledge can be characterized as contradicting the typical complaint in vocational education that theory and practice often remain separate domains for students and interns (Guile 2006; Tynjälä 2008). The case study also shows that internship with good preparation and supervision can provide a powerful learning environment in which even a less motivated student can be enthused for a research project involving statistical inference. The involvement in authentic decision-making seems to have been crucial.

The third point to note is that an inferentialist perspective may be useful beyond the vocational one. Coordination and related concepts, such as coherence and integration, are major themes in general education as well (Linn 2006; Nilsson 2009), for example when promoting coherence between mathematics and science (Dierdorp et al. 2014). The literature on coherence and integration (e.g., Berlin and Lee 2005) pays attention to curricular issues, but little attention to what may help elements to cohere or become integrated at a micro-level. An inferentialist view points to the importance of inferential relations and reasons as the glue. Inferentialism, of course, does not in itself provide guidelines on which inferential relations task designers should focus, but can presumably help to frame theory on task design. For example, Ainley et al. (2006) argue that purpose and utility are important foci in task design. A task needs to have a purpose from a student perspective (a reason to come to action) and what students learn needs to have utility; it has to be useful for solving a problem they care about. From an inferentialist perspective, purpose and utility are among the many possible reasons in larger webs. In Sam's case, the purpose was clear: to test whether it was necessary to bring particular blood samples to the laboratory by foot. The utility of what he learned was also clear: chemical and statistical knowledge helped him draw a valid conclusion that the hospital cared about.

Another area where inferentialism may help to make progress is in the realm of webbing and situated abstraction. These are potentially useful concepts, but have hardly been further developed since Noss and Hoyles (1996) and Pratt and Noss (2002). In particular, the nature of coordination and how exactly reason and action are webbed has remained undertheorized. In analyzing how learners construct meaning, Noss and Hoyles (1996) argued for a perspective of "learning as the construction of a *web* of connections—between classes of problems, mathematical objects and relationships, 'real' entities and personal situation-specific experiences" (p. 105, emphasis in original). Their idea of *webbing* is domain contingent and conveys a structure that learners can choose to step onto, rely on, and/or reconstruct for support as they wrestle with meaning-making (cf. Bakker 2014a). Noss and Hoyles (1996) see webbing and situated abstraction as complementary, challenging the separation between what is learnt and how it is learnt (p. 120), implying that situations and learners "co-produce knowledge through activity" (citing Brown et al. 1989, p. 32). Situations (whether they be classroom activities or laboratory contexts) generate linguistic and conceptual resources that extend and point to structures beyond specific activities. Abstraction becomes a resource for activity when new meaning is created and learners (or laboratory technicians) exploit the invariances they experience to achieve a goal.

Like the technological environments in which Noss and Hoyles applied their idea of webbing, the authentic context of the laboratory acts as an action/linguistic framework to support the creation of meaning. Rather than ignore the agency of the learner, the situatedness of the context, and the way that the setting could create structures, Noss and Hoyle's (1996) perspective of webbing challenges "the notion that mathematics achieves its power by suppression rather than extension of meanings, and that the very existence of meaning mitigates against generality" (p. 130). Rather than provoke a dichotomy between situation-specific knowing and generality, webbing blurs this division by acknowledging meaning as being reshaped and recreated in action. Our study has shown how the coordination of actions and reasons is one way in which such webbing can take place.

More generally, we think that Brandom's inferentialist ideas help us study knowledge and action in relation to each other while avoiding a dichotomy between them (cf. Heusdens et al. 2016). Other dichotomies that can be avoided with the help of inferentialism are individual versus collective cognition (Schacht and Hußmann 2015) and statistical versus contextual knowledge (Bakker et al. 2008).

Despite Brandom's somewhat rationalist language, his ideas fit well with the Vygotskyan emphasis on not separating cognition from activity and affect (Vygotsky 1986). Our example on Sam's pride and the happiness expressed by the hospital supervisor point to the possibility that emotions can be part of webs of reasons and actions (an emotion can be a reason for an action, but an action can also be the reason for an emotion). More generally, Derry (2008, 2013) has argued that Brandom's ideas are compatible with Vygotsky's (which were also influenced by Hegel), which makes Brandom's inferentialism interesting to sociocultural researchers. At the same time, Brandom gives a careful account of concepts and concept use, which is not so common in sociocultural research with its focus on practice and activity (as observed by Guile and Young 2003). Analyses of practices or activity systems common in sociocultural traditions tend to focus on what is done for a general objective. Such a focus does not necessarily give much insight into the fine-grained inferential structure of practices or how actions, knowledge, and emotion can be interwoven. Yet such fine-grained analysis may be useful for designing learning environments, not only in vocational education where actions and practical knowledge are evidently important (Bakker and Akkerman 2014), but also in general education that aims to take coordination (coherence and integration) seriously and seeks for holistic development of students' knowledge, skills, and attitudes.

Acknowledgements This research is made possible with the financial support of a grant from the Educational and Learning Sciences Utrecht to Arthur Bakker and a grant DP120100690 by the Australian Research Council to the three authors. We thank Jean Lave for encouraging us to elaborate the idea of conceptualizing reasons as being relational. We thank participants of SRTL-8 and of the Inferentialism in Statistics and Mathematics Education group (in particular Stephan Hußmann) for their helpful feedback on earlier versions of this article.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

Ainley, J., Pratt, D., & Hansen, A. (2006). Connecting engagement and focus in pedagogic task design. *British Educational Research Journal, 32*(1), 23–38.

Arnold, P., Budgett, S., & Pfannkuch, M. (2013). Experiment-to-causation inference: the emergence of new considerations regarding uncertainty. *Proceedings of the Eighth International Research Forum on Statistical Thinking, Reasoning and Literacy* (srtl-8) (pp. 119–146). Two Harbors, MN: The University of Minnesota.

Bakker, A. (2014a). Characterising and developing vocational mathematical knowledge. *Educational Studies in Mathematics, 86*(2), 151–156.

Bakker. (2014b). Implications of technology on what students need to know about statistics. In T. Wassong, D. Frischemeier, P. R. Fischer, R. Hochmuth, & P. Bender (Eds.), *Mit Werkzeugen Mathematik und*

Stochastik lernen—using tools for learning mathematics and statistics (pp. 143–152). Wiesbaden: Springer.

Bakker, A., & Akkerman, S.F. (2014). A boundary-crossing approach to support students' integration of statistical and work-related knowledge. *Educational Studies in Mathematics*. http://link.springer.com/article/10.1007%2Fs10649-013-9517-z.

Bakker, A., & Derry, J. (2011). Lessons from inferentialism for statistics education. *Mathematical Thinking and Learning, 13*, 5–26.

Bakker, A., Kent, P., Derry, J., Noss, R., & Hoyles, C. (2008). Statistical inference at work: the case of statistical process control. *Statistics Education Research Journal, 7*(2), 130–145.

Bakker, A., Wijers, M., Jonker, V., & Akkerman, S. F. (2011). The use, nature and purposes of measurement in intermediate-level occupations. *ZDM—The International Journal on Mathematics Education, 43*(5), 737–746.

Ben-Zvi, D. (2006). Scaffolding students' informal inference and argumentation. In A. Rossman, & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics* (CD-ROM), Salvador, Bahia, Brazil, July, 2006. Voorburg, The Netherlands: International Statistical Institute.

Ben-Zvi, D., & Aridor, K. (2016). Children's wonder how to wander between data and context. In D. Ben-Zvi & K. Makar (Eds.), *The teaching and learning of statistics: international perspectives* (pp. 25–36). New York: Springer.

Berlin, D. F., & Lee, H. (2005). Integrating science and mathematics education: historical analysis. *School Science and Mathematics, 1*, 15–24.

Brandom, R. B. (1994). *Making it explicit. reasoning, representing, and discursive commitment*. Cambridge: Harvard University Press.

Brandom, R. B. (2000). *Articulating reasons: an introduction to inferentialism*. Cambridge: Harvard University Press.

Brandom, R. B. (2002). Non-inferential knowledge, perceptual experience, and secondary qualities: placing Mcdowell's empiricism. In N. H. Smith (Ed.), *Reading Mcdowell: on mind and world* (pp. 92–105). Abingdon, UK: Routledge.

Brandom, R.B. (2008). *Between saying and doing: towards an analytic pragmatism*. Oxford University Press.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32–42.

Derry, J. (2008). Abstract rationality in education: from Vygotsky to Brandom. *Studies in Philosophy and Education, 27*, 49–62.

Derry, J. (2013). *Vygotsky: Philosophy and education*. Oxford: Wiley-Blackwell.

Dewey, J. (1929/2008). *The quest for certainty: a study of the relation of knowledge and action* (Gifford lectures 1929). New York: Minton, Balch & Company.

Dierdorp, A., Bakker, A., Eijkelhof, H. M. C., & van Maanen, J. A. (2011). Authentic practices as contexts for learning to draw inferences beyond correlated data. *Mathematical Thinking and Learning, 13*, 132–151. doi:10.1080/10986065.2011.538294.

Dierdorp, A., Bakker, A., van Maanen, J.A., & Eijkelhof, H.M.C. (2014). Meaningful statistics in professional practices as a bridge between mathematics and science: an evaluation of a design research project. *International Journal of STEM Education*, *1*(9).

Ellis, P. D. (2010). The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results. Cambridge University Press.

Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: connecting research and teaching practice*. New York: Springer.

Guile, D. (2006). Learning across sites. *Educational Philosophy and Theory, 38*, 251–268.

Guile, D., & Young, M. (2003). Transfer and transition in vocational education: Some theoretical considerations. In T. Tuomi-Gröhn & Y. Engeström (Eds.), *Between school and work: new perspectives on transfer and boundary-crossing* (pp. 19–38). Amsterdam: Pergamon.

Heinicke, S., & Heering, P. (2013). Discovering randomness, recovering expertise: the different approaches to the quality in measurement of Coulomb and Gauss and of today's students. *Science & Education, 22*, 483–503. doi:10.1007/s11191-011-9430-8.

Heusdens, W., Bakker, A., Baartman, L., & De Bruijn, E. (2016). Contextualising vocational knowledge: a theoretical framework and illustrations from culinary education. *Vocations and Learning, 9*(2), 151–165.

Hußmann, S. & Schacht, F. (2009). Toward an inferential approach analyzing concept formation and language processes. *Proceedings of Cerme 6*, January 28th–February 1st 2009, Lyon.

Linn, M. C. (2006). The knowledge integration perspective on learning and instruction. In K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences*. Cambridge: MA. Cambridge University Press.

Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal, 8*(1), 82–105.

Makar, K., Bakker, A., & Ben-Zvi, D. (2011). The reasoning behind informal statistical inference. *Mathematical Thinking and Learning, 13*, 152–173.

McDowell, J. (1996). *Mind and world (with a new introduction by the author)*. Cambridge: Harvard University Press.

Moore, D. S. (1995). *The basic practice of statistics*. New York: W.H. Freeman and Company.

Nilsson, P. (2009). Conceptual variation and coordination in probability reasoning. *The Journal of Mathematical Behavior, 28*(4), 247-261.

Noss, R., & Hoyles, C. (1996). *Windows on mathematical meanings: learning cultures and computers*. Dordrecht: Kluwer Academic Publishers.

Pfannkuch, M. (2006). Informal inferential reasoning. In A. Rossman & B. Chance (Eds.), *proceedings of the Seventh International Conference on Teaching Statistics. Working cooperatively in statistics education*, Salvador, Brazil, July, 2006 [cd-rom.] Voorburg, The Netherlands: international statistical institute. http://iase-web.org/documents/papers/icots7/6a2_pfan.pdf.

Piaget, J. (1970/2013). *Principles of genetic epistemology: selected works* (Vol. 7; Translated by W. Mays). London: Routledge.

Pratt, D. (2012). *Making mathematics phenomenal (inaugural professorial lecture)*. London: Institute of Education.

Pratt, D., & Noss, R. (2002). The microevolution of mathematical knowledge: the case of randomness. *The Journal of the Learning Sciences, 11*(4), 453–488.

Roth, W.-M. (2014). Rules of bending, bending the rules: the geometry of conduit bending in college and workplace. *Educational Studies in Mathematics, 86*(2), 177–192. doi:10.1007/s10649-011-9376-4.

Sanders, T. J. M., Spooren, W. P. M., & Noordman, L. G. M. (1992). Toward a taxonomy of coherence relations. *Discourse Processes, 15*, 1–35.

Schacht, F., & Hußmann, S. (2015). Between the social and the individual: reconfiguring a familiar relation. *Philosophy of Mathematics Education Journal, 29*, 1–26.

Schindler, M., & Hußmann, S. (2013). About students' individual concepts of negative integer—in terms of the order relation. In the *Eighth Congress of European Research in Mathematics Education* (CERME-8) *Proceedings* (February 2013, Manavgat-Side, Antalya–Turkey). http://cerme8.metu.edu.tr/wgpapers/WG2/WG2_Schindler.pdf.

Sellars, W. (1956/1997). *Empiricism and the philosophy of mind. With an introduction by Richard Rorty and a study guide by Robert Brandom*. Cambridge, MA: Harvard University Press.

Tynjälä, P. (2008). Perspectives into learning at the workplace. *Educational Research Review, 3*, 130–154.

Vygotsky, L. S. (1986). *Thought and language (newly revised and edited by A. Kozulin)*. Cambridge, MA: MIT Press.

Wedege, T. (1999). To know—or not to know—mathematics, that is a question of context. *Educational Studies in Mathematics, 39*, 205–227.

Wild, C. J., Pfannkuch, M., Regan, M., & Horton, N. (2011). Towards more accessible conceptions of statistical inference (with discussion). *Journal of the Royal Statistical Society: Series A (Statistics in Society), 174*(2), 247–295.

Womack, J. P., Jones, D. T., & Roos, D. (1991). *The machine that changed the world: the story of lean production*. New York: Harper Perennial.

Zieffler, A., Garfield, J., DelMas, R., & Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal, 7*(2), 40–58.