7 Building Trust in and With Human Computation[1]

"[W]e have to trust the whole citizen science stuff" (fieldnote Jul. 27, 2021), explained researcher Anna during the first laboratory meeting I attended for my ethnographic fieldwork in Ithaca in 2021. In this meeting, the laboratory discussed recent improvements in the data pipeline that connects the laboratory's Alzheimer's disease research with the online game Stall Catchers. Although this statement struck me as remarkable at the time, I had not considered "trust" an empirical category of interest or an analytical concept for my research on HC systems. However, throughout my first three-month stay in Ithaca to learn about the researchers, developers, and designers' perspectives on their joint endeavor, Stall Catchers, the notion of trust came up repeatedly in conversations with laboratory members. Moreover, trust also emerged as a critical aspect of legitimizing the HC-based CS approach to knowledge production pursued at the Human Computation Institute and its mission to "engineer sustainable participatory systems that have a profound impact on health, humanitarian, and educational outcomes" (Human Computation Institute, n.d.).

What does it mean for biomedical researchers to "trust the whole citizen science stuff"? Trust, as I will show in this chapter, plays a crucial role in the formation and maintenance of HC-based CS assemblages. By analyzing not only the articulations but also the practices of researchers in human–technology relations, it becomes clear that trusting the "citizen science stuff" is not only a social relation but also includes sociomaterial[2] practices that continuously reestablish trust. The reappearance of the notion of trust in the field brought me a new analytical perspective on how these systems are continuously becoming in the interplay of the different human and nonhuman actors involved. I consider trust as reterritorialization processes that bring together and align various elements and relations to (re)configure these sociotechnical systems. Trust as a

1 This chapter builds on ideas presented in the talk "Between means and ends: Data infrastructures in biomedical research" at the RAIMed conference 2022 (Vepřek 2022a), and a poster presentation at the Spring School of the *Transformations in European Societies* Ph.D. program in Murcia, Spain, on March 25, 2022.

2 In this chapter, I use the term socio*material* practice, which includes socio*technical* practices, because the point I would like to make is not specific to technologies but refers more broadly to how trust is built in relations of humans and materialities.

sociomaterial practice is at the core of HC development as it emerges and needs to be adapted alongside the intraverting relations in HC-based CS due to new changes and shifts unsettling established trust-building mechanisms.[3] Trust, as I understand it in this chapter, is itself created within human–technology relations.

To unfold this perspective, I revisit some of the examples discussed previously through the lens of trust as a sociomaterial practice, while also discussing some new instances of HC-based CS and its human–technology relations. The chapter is organized as follows. First, I briefly discuss trust as an analytical concept before delving into various trusting practices and relations in the example of Stall Catchers. I then analyze trust within HC-based collaborations from the researchers' point of view and their understanding of scientific confidence. HC itself plays an essential role in building trust in the CS approach. At the same time, trust in HC systems must be programmed algorithmically, as I discuss from the perspective of the Human Computation Institute's team. Next, I turn to the participants' perspective and their trust in the researchers and the Stall Catchers team. Before concluding this chapter, I briefly discuss the question of proprietary software and trust. The aim of this chapter is not to disregard trust as a social relation but to open up the concept of trust to sociomaterial practice to gain a deeper understanding of human–technology relations in HC-based CS. The understanding of trust suggested in this chapter, thus, does not seek to define trust once and for all—which is an impossible undertaking anyway, as human computer interaction researcher Richard Harper shows in the edited volume *Trust, Computing, and Society* (2014a)—but rather to provide a valuable concept for analyzing the field at hand. While trust emerged from the field research on Stall Catchers and this chapter is, therefore, largely based on this example, it also draws on Foldit in some places as a comparative study to further reflect on the larger context of trust in the field.

Trust as an Analytical Concept

Trust, as a sociological concept, has received considerable attention.[4] In 2000, sociologist Piotr Sztompka described a "new wave of sociological interest in trust" (2000, 14),[5] which he attributes not only to a variety of reasons, such as the growing complexity and

3 My interest is in how trust is being built in HC-based CS. This exploration necessarily includes considering instances of lack of trust or mistrust. Anthropologist Florian Mühlfried suggests that mistrust, compared to trust, remains an understudied phenomenon (2018, 7). Mühlfried defines mistrust, following sociologist Niklas Luhmann, as a way of reducing the complexity of the world. Mistrust seeks out "defensive arrangements" (Luhmann 2014, 1), that, according to Mühlfried, are "ways to spread risks and weaken dependencies" (2018, 11). However, I do not delve further into conceptualizations of mistrust or distrust (on the relation between mistrust and distrust, see, for example, Carey [2017, 8]; *cf.* Mühlfried [2018]), but refer to the terms as they are used empirically by my research partners.

4 It is not my aim here to provide a comprehensive overview of the existing literature on trust. To point to further work beyond that discussed, see, for example, Garfinkel (1963), Braithwaite and Levi (1998), Gambetta (1988b), Apelt (1999), Endreß (2012), Schilcher, Will-Zocholl, and Ziegler (2012), Weichselbraun, Galvin, and McKay (2023).

5 For a review of sociological theories on trust in the twentieth century, see Sztompka (2000).

interdependence of the world and human relations but also to the indeterminacy of social roles or the increasing opaqueness of institutions and technological systems. Scholars, such as anthropologist Alberto Corsín Jiménez, have also observed a "crisis of trust" (2011) over the last few centuries. This crisis manifests itself in a decrease of trust not only in the state but also in institutions, and, as can be observed in the climate crisis and during the COVID-19 pandemic, in science. Contrary to earlier psychological understandings of trust as a personal attitude, according to Sztompka, trust is now conceived as "the trait of interpersonal relations, the feature of the socio-individual field in which people operate, the cultural resource utilized by individuals in their actions" (2000, 14). The interactive and interpersonal nature of trust (Weingardt 2011, 9) is an important aspect in the literature on trust, which is commonly understood as (a) social relations(hips) (e.g., Luhmann 1988; 2014; Hardin 2006; Weingardt 2011).[6] According to Sztompka, trust is something inherently human—it cannot be bound to natural phenomena (2000, 20). What characterizes trust is its future orientation and its association with the uncertainty and uncontrollability of the future (Sztompka 2000, 20). According to sociologist Niklas Luhmann, trust serves to reduce the complexity of the world and is both a risky investment and a "solution for specific problems of risk" (1988, 95).[7] However, trust does not refer to *any* future action but to those that shape our present decisions (Gambetta 1988a, 218–219). Trust, therefore, fills the gap between risky and unpredictable futures and the need to make decisions and take action. Sociologist Heinz Bude, following Luhmann, has aptly summarized the essence of trust:

> Trust opens up perspectives and enables action in complex and complicated situations. Wittgenstein provided a philosophical explanation for this when he explained how, through trust, the reason of life overrules the unreason of doubt. This is, as Niklas Luhmann has poignantly demonstrated, the logic of a risky advance, which is only ever rational in retrospect. From the feeling of trust, one dares to take the leap into the dark, which overcomes the hiatus between justification and decision. (Bude 2010, 11)

This sense of trust, sociologist Anthony Giddens contends, is not limited to social relationships or individuals but can also be extended to abstract systems (1990).[8] Giddens employs the term "abstract systems" to refer to symbolic tokens or expert systems that play a significant role in structuring life in the modern era. The subtle difference between these forms of trust is illustrated by the example of money: "it is money as such which is trusted, not only, or even primarily, the persons with whom particular transactions are carried out" (Giddens 1990, 26). As this quote indicates, however, trust in persons is

6 Some authors divide trust as social relations further into trust in oneself, trust related to certain networks of the lifeworld of which one is part, and trust in systems and institutions (Bude 2010, 13).

7 According to Luhmann (2014), distrust is the functional equivalent to trust. Therefore, they both function in the same way.

8 Sociologist Martin Endreß defines trust as a multidimensional phenomenon by considering three different modes of trust: reflexive, habitual, and pre-reflexive functioning mode (2012).

still, to some extent, involved in trust in systems regarding their "*proper* working" (Giddens 1990, 34, emphasis i.o.). Gidden's distinction is valuable for the following analysis because trust in the field includes the social dimension but also goes beyond it. However, like the sociological theories discussed previously, Giddens did not pay particular attention to the materialities (or sociomaterialities) for building trust, which play an essential role in understanding trust in sociotechnical systems, such as HC-based CS projects. In fact, most theories of trust in sociological and social theory, despite varying definitions of trust and related concepts, such as confidence (e.g., Luhmann 1988; Giddens 1990; Seligman 2000),[9] share the understanding of trust as a cognitive, social, and only human phenomenon. Yet, trust, as it was expressed and performed in the field, cannot be captured by a cognitive understanding of trust alone because, as Corsín Jiménez aptly describes, trust is "also distributed in a variety of human and nonhuman forms; it is as much a cognitive category as it is a material one; indeed, it belongs to the realm of the intersubjective in as much as it belongs to the interobjective" (2011, 179). Although Corsín Jiménez explicitly points out that his work does not aim to contribute to the existing literature on trust, it forms a fruitful basis for the following reflections since, as I will show below, trust is built *along* and *with* technology in the field of HC-based CS projects (Ingold 2007).[10] In postphenomenological terms, it is mediated by technology. Researchers in the laboratory are continuously working on establishing trust with technologies precisely because "trust [in] the whole citizen science stuff" (fieldnote Jul. 27, 2021) cannot be based entirely on social relations. In HC-based CS projects, scientists, developers, and participants enter new collaborations with each other and other actors, thereby introducing new relations and practices previously unknown to the individual parties and partly requiring the reevaluation of established processes and practices previously used to build and maintain trust.

---

9 Luhmann and other scholars have distinguished trust from other concepts, most prominently from the concept of confidence. According to Luhmann, trust and confidence share that they are both "modes of asserting expectations" (1988, 99), whereby these expectations can turn into disappointments (1988, 97). However, while confidence is the case when one has no other choice than to be confident in something, trust is connected to one's active previous engagement. Trust implies risk in a particular situation in which one decides to act in one way or the other (Luhmann 1988, 97–98). However, relations of trust or confidence can transform into the other. The distinction between trust and confidence is also common in everyday life, as the proverbs such as "trust, but verify" or the German form *Vertrauen ist gut, Kontrolle ist besser* demonstrate (*cf.* Seligman 2000, 17). While religious studies scholar Adam Seligman agrees on the distinction between trust and confidence, he argues that "trust" in its current understanding "as a solution to a particular type of risk" (2000, 7–8) is a phenomenon specific to modernity and not universally transferable to social organization in general (2000, 6). Confidence, by contrast, is required for any social organization to work. On the other hand, Giddens defines trust as "a particular type of confidence" (1990, 32) and, thus, not as something different from confidence. Similarly, during my field research, I could not observe a clear distinction between confidence and trust as suggested by Luhmann and Seligman. Instead, and as I will show in this chapter, trust, confidence, and control merged smoothly into each other and were sometimes interwoven to an extent that it was not possible to clearly differentiate between them.

10 Pink et al. have also shown how trust can be built through "familiar technologies" (2018, 11), such as paper documentation.

I analyze human–technology relations in practice to grasp how trust emerges and is maintained in collaboration and knowledge production in my case study. In this sense, and building upon recent anthropological approaches to trust (e.g., Pink, Lanzeni, and Horst 2018; Pink 2021; 2022; 2023; Weichselbraun, Galvin, and McKay 2023), I explore trust as a sociomaterial practice that is distributed between human and nonhuman actors, such as software tools and data flows. While the understanding of trust as a purely cognitive phenomenon seems too narrow for the analysis of HC-based CS (and, I think, sociotechnological or technologically mediated lifeworlds in general), the framework of trust as sociomaterial practice includes trust as social relations and opens it to human–technology relations in practice.[11] In this way, it also allows one to capture the understandings of trust that appear in the field studied itself without requiring a strict definition of the boundaries of trust—such as defining in advance where trust turns into confidence or belief—and, hence, remaining open to the meanings and practices that unfold in HC-based CS. Focusing on practices, moreover, follows a "processual theory of trust" (Pink, Lanzeni, and Horst 2018, 11), as suggested by Sarah Pink, anthropologist Deborah Lanzeni, and sociocultural anthropologist Heather Horst, which "maps out how people cope with the inevitable uncertainty and contingency of the emergent circumstances of everyday life" (2018, 11–12). Trust as sociomaterial practice, thus, resembles Pink's approach to "everyday trust," which understands trust as an anticipatory concept that "involves 'a sensory experience of feeling or disposition towards something' rather than an explicit cognitive decision made in relation to a specific technology" (Pink, Lanzeni, and Horst 2018; cited in Pink 2022, 47). In what follows, however, I aim to focus not so much on the feeling (as a sensory experience) of trust as a result of specific configurations (Pink 2021, 193) but on the trust-generating practices themselves unfolding in relations between humans and technology that are always situated in everyday and historical contexts. By focusing on practice, it also connects to approaches to trust that see trust as a "*doing* rather than a fixed point" (Garfinkel 1967; cited in Harper 2014b, 324, emphasis i.o.; *cf.* Watson 2014). Trust, then, is "ephemeral, emergent, contingent and shifting" (Pink 2023, 29) and, as such, is not independent of mistrust but interdependent with it (Pink 2023, 38; Mühlfried 2018, 11).

In the following subchapters, I will jump back and forth between both trust as an empirical category and an analytical concept to capture the different layers of trust involved in HC-based CS projects and foster the dialogue between them. Here, a basic tension between the analytical and empirical term may always remain, but I hope to turn it into a productive one.

11 Trust as a sociomaterial practice, thus, does not form an alternative or contrast to trust as social relations but, instead, a broader understanding that encompasses both human–human, human–technology, technology–human–technology, etc. relations.

Trust as a Sociomaterial Practice

"We Have to Trust the Whole Citizen Science Stuff"

Science is an inherently unpredictable adventure, and it is part of doing science to "wrestle with the unknown" (Schaffer, Dec. 7, 2021). Research pushes the boundaries of knowledge; therefore, it must deal with scientific uncertainty. Schaffer explained in our interview that "scientific uncertainty is something that […] we naturally deal with" (Dec. 7, 2021). While this applies to research in general, researchers at the laboratory studied described Alzheimer's disease as a particularly complicated research subject. Very little is known about the mechanisms of the disease, and multiple factors seem to influence its onset and progression. A laboratory member summarized the challenges they face in studying Alzheimer's disease: "[w]hat a disaster of a disease" (fieldnote Sept. 07, 2021). Following Lock (2013), the laboratory's approach to Alzheimer's disease can be described with the term "localization theory" (2013, 5). Lock identifies two approaches to Alzheimer's disease research. Regarding localization theory, Lock writes, "neuropathological changes in the brain are assumed to be causal of specific behavioral changes in persons" (2013, 5). The second approach is the "entanglement" theory of dementia, in which advocates "favor theories about the way in which mind, persons, life events, aging, and environments interact to precipitate neurological and behavioral transformations that are pathological" (Lock 2013, 5). While localization theory was dominant in the twentieth century, today, there is a growing awareness of entanglements of the environment, the mind, and the body (Lock 2013, 5), as is evident in the laboratory member's quote about the disastrous disease. Despite the researchers' awareness of different entanglements at the studied laboratory, most of the experiments related to Alzheimer's disease revolved around the question of how stalled blood vessels could be resolved and blood flow restored. In our conversation in late October 2021, researcher Jada reflected on this research focus as a very small detail of Alzheimer's disease: "[E]verything we do is a minor aspect to the very complicated disease process that […] may or may not turn out to be—even in years from now—to be really important or not. We just don't know at this point. So, it looks like now […] it's likely going to be important, but we don't know that" (Oct. 27, 2021). Uncertainty about the future success of their current scientific efforts was part of the researchers' daily experience.

Apart from the general scientific uncertainties, there are other uncertainties specific to the research process of Alzheimer's disease research and the laboratory's collaboration with Stall Catchers. Research on Alzheimer's disease in the laboratory was marked on a daily basis by the unpredictability of experiments and their outcomes due to contingencies and the unruliness of the nonhuman actors involved. This included, for example, the material used for the chronic cranial windows, which were installed during surgery to allow subsequent *in vivo* imaging of the mouse brains, or the activity and life cycles of mice, which did not always align with the experimental procedures. How good will the image quality of an individual imaging session be? How long will a mouse survive with a new treatment? Conducting scientific research in this area necessarily included recognizing and accepting the unpredictability of experiments and the potential challenges that can arise during a research project.

The development of Stall Catchers and its introduction into the routines of researchers at the laboratory added even more previously unknown uncertainties. This was due in part to the nature of Stall Catchers as "a cutting-edge experiment," Schaffer explained (Dec. 7, 2021), a not-yet-well-established approach to analyzing scientific research data for which there were no solutions and which introduced questions about data quality and impact on the research at the laboratory which could not be anticipated at the beginning. Furthermore, unpredictable factors and, in particular, the large amount of work laboratory members had to invest in Stall Catchers without receiving any immediate benefit from the platform made it difficult to convince them of the value of the project, "[a]nd it took a little bit of pushing from Chris [Schaffer] to get the lab and everybody sort of buy-in on—because it took quite concerted effort without seeing payback for a couple of months," explained Nishimura (Dec. 7, 2021).

Although adapting work routines and methods to new technologies and tools was described as difficult for many laboratory members in general, researcher Isabel said it was especially difficult to convince some laboratory members of new steps and tools related to the CS game. She explained that this was also partly due to the distance between the laboratory and its collaborators: "Stall Catchers is a harder thing to do, it's a longer turnaround, and there is not as much of a direct feedback […] [T]here is this lab, and then Stall Catchers and the Human Computation Institute is over there somewhere" (Isabel, Oct. 14, 2021).

The long turnaround and the novelty of the Stall Catchers approach added to the "usual" uncertainty researchers were used to. Moreover, and this will be the focus of the following pages, they could not rely on established practices and their familiar routines to work against uncertainties. Doing science was defined not only by the inherent or "natural" uncertainty of scientific research but also by the "commitment to […] getting it right" (Nishimura, Dec. 7, 2021). Nishimura explained that her motivation to do science

> is that I really want […] the answers that we find to be correct. And so in the field of Alzheimer's disease, it's really important to me that even if we're answering a small question or a big question, doesn't matter, and it does need to be correct, and […] something that I really worry about is getting something wrong that is eventually going to feed into a drug that doesn't work or an idea that goes the wrong way or something like that. So that is something that I really do worry about and care about. And I think a lot of people in the lab are also motivated by what I call as a commitment to; I don't know if you want to personify like truth or something […] that you really commit to. But getting it right. (Nishimura, Dec. 7, 2021)

The drive to be 100 percent certain of a new finding, to eliminate all uncertainty, was not a contradiction to the uncertainty that defines scientific inquiry but stood next to it. Scientific work unfolds in this space between uncertainty and "getting it right," and this is where the notion of trust comes into play.

During my first research visit to Ithaca in 2021, trust came up in most of my conversations with laboratory members when they discussed current problems in Alzheimer's disease research involving Stall Catchers. The term was most often used to describe what was not working well and why they could not always use data analyzed by Stall Catchers.

When I mentioned this observation in my conversation with the PIs, they explained that, when talking about "trust," they were not referring to trust "in people" but trust as "scientific confidence." It was about trust "in the process and in the system" (Schaffer, Dec. 7, 2021). This process- and system-oriented understanding of trust appears to be typical of HC systems as a whole, where the crowd is at a conceptual level, an anonymous, depersonalized group of individuals (see Chapter 4) and is also reflected in the development of these systems, as will be further explained later in this chapter. Trust has to be built differently when there is no person on the other side to address directly.

When laboratory members talked about ongoing problems with research building on Stall Catchers, they consistently emphasized that these were not Stall Catchers' fault. Researcher Emily stressed that "it's not their [Stall Catchers'] fault, it's, I think it's all on our side" (Sept. 8, 2021). The problems were rooted in the data pipeline, in the preprocessing of the research data to be sent to Stall Catchers. More specifically, as I discussed from the perspective of infrastructuring, members explained that the problem was that they could not trust the "absolute number" (Schaffer, Dec. 7, 2021) of stalls that resulted from the crowd annotations—not because they did not trust the crowd's accuracy in finding stalls, but because the preprocessing did not output the exact number of capillaries. Capillaries got lost in the data pipeline:

> [T]he holdup really is not so much […] whether a vessel is stalled or not. Stall Catchers is really good at that. We're kind of getting stuck because we need to know the total vessel count and that's where … it's … getting a little more frustrating because Stall Catchers can, if we send a vessel that's not a vessel, Stall Catchers will mark it as not a vessel […] most of the time. But that's also not fun for them if there's a whole bunch of garbage that we're sending […]. So that's kind of where we're at right now. […] I've been trying to get the data nicer for their sake and our sake. Using computers, not us, because initially, it was on us to do it manually. (Leander, Sept. 22, 2021)

This problem only became clear after the introduction and initial development of the CS project. Previously, one of the most important questions was whether a vessel was stalled. The question of how many vessels there are in the first place was introduced with Stall Catchers and along the intraverting researcher–technology relations in the pipeline. It had not arisen earlier because the manual research data analysis had been performed on the "raw" image data, where the vessel count could be relied upon. However, with the current state of the data pipeline and the preprocessing of the data, "the data we are sending, you had a whole ton of non-vessels and a whole bunch of broken-up vessels and things like that. So, we couldn't really trust that aspect of it" (Isabel, Oct. 14, 2021). Previous trust-building relations were disrupted with the introduction of Stall Catchers and the laboratory's new collaboration with the Human Computation Institute and Stall Catchers participants, as well as the new infrastructure; trust or scientific confidence had to be rebuilt.

In the following pages, I will show and discuss how trust had been established and maintained prior to the introduction of Stall Catchers and how it has evolved since the introduction of the CS platform in various sociomaterial practices.

In order to better understand the differences between the manual analysis of research data conducted in the laboratory and the analysis conducted by the crowd in Stall Catchers, the following excerpt from a field note from one of my laboratory visits in late September 2021 provides insights into the manual annotation of stalls.

> "I cannot promise that it will be exciting," explains James while plugging in a hard disk with the image stacks. He opens the folder containing another 18 folders with data of a different image session of a specific mouse, each including one image stack. James has been working on the analysis of this specific dataset for about two and a half months now. He explains that they have an extra spreadsheet with the depths of each image stack, so he knows how deep the imaging process went and how many frames he has to analyze. James picks a yellow post-it and writes down 10/20/30/40/50/60/ … /470/480, each number representing a projection of ten slices. He will look at one projection, e.g., ten images/slices, at a time and cross out those numbers that he's already analyzed to ensure that he does not miss a slice.
>
> Before analyzing the images, James must first adjust the image program settings to fit the to-be-analyzed image stack. For example, he has to set the number of color channels [with two-photon microscopy, different color channels can be imaged simultaneously], and the depth of the image stack. James, furthermore, adjusts the brightness of the image to a contrast that is more pleasant for his eyes and sets the last channel to gray, which is easier for him to see. James keeps the little menu for the brightness setting open during the analysis session. "And from here, it's the same over and over again," he explains.
>
> As James progresses through the image stack, he frequently adjusts the brightness level to compensate for the increased graininess of the pictures at deeper levels. This helps to improve visibility and to discern finer details in the images. During the analysis, there are windows for each color channel with images from one image stack. James begins with the analysis of the first ten slices. He opens an additional window with the projection of the slices and looks for something "suspicious," a suspicious area or spot that could be a stall. If James finds something, "I keep an eye on it and go fast" through the image stack. Therefore, he switches to the combined channel, and by navigating with the right and left arrow buttons, he goes through the image stack, looking for stalls at this specific position. If he is not sure, he switches to one of the other channels depicting only specific structures or molecules. A stall, here, refers to a blockage in a capillary that doesn't start flowing within five slices. Whenever James finds one, he marks it in the channel and saves the image with the stall in an extra stalls folder.
>
> He explains that "you do everything you can to make it go faster." So, instead of analyzing one projection at a time, James opens two projections and works on them in parallel. Additionally, he uses hotkeys to speed up the clicking procedures. Not all experimentalists analyze projections of ten; others prefer to analyze projections of 20, and they all have their own pace and practices. After around 30 minutes, James completed today's first image stack. (fieldnote, Sept. 27, 2021)

Manually annotating images, as described in the fieldnote, can be summarized as dividing an image stack into projections of a few slices each, which are then compared with the scrollable image stack displaying certain channels in separate windows. Here, "scientific confidence" emerges in the manual annotation of stalls by relying not only on

the experimenter's own eyes and practice to identify stalls but also through the "raw" image stacks being analyzed, which have not been manipulated by any processing algorithms—even though the data has already gone through different steps of translation (Callon 1984; Latour 1999) from excited photons captured by the microscope's objective to scrollable stacks of TIFF-format files depicting two-dimensional image slices. Interestingly, trust only became a concern after image generation. Researchers did not describe any gaps in the microscopic imaging process and the process of generating raw imaging data. To further reduce bias and error, the image stacks were analyzed by at least two laboratory members and were anonymized to "blind" the experimenters. They were not to know whether they were analyzing the data from an Alzheimer's mouse with a particular treatment or a control mouse.

Now, with the introduction of collaboration with the HC-based CS project and, thus, with the introduction of new data infrastructures that made the crowd analysis possible in the first place, these established trust-building practices no longer worked. As will be described below, this first created a lack of trust, requiring it to be built differently through new human–technology relations. Two different moments, which overlap in practice, can be distinguished analytically here: trust in the now outsourced analysis had to be (re)established and trust in the new automated preprocessing and manipulation of data—in short, the data pipeline—had to be built.

At the time of my fieldwork at the laboratory, some experimentalists still preferred to manually annotate stalls rather than "use" (James, Sept. 27, 2021) Stall Catchers. The use of Stall Catchers required researchers to rely on an anonymous crowd of people with no formal training to analyze the data. There remained a perceived gap between the researchers' own analysis and the crowd. One reason for this can be linked to a question of routine and the pleasure of working and engaging with the research data. Researcher Anna described the difficulty of handing over tasks to someone else in general and, as she described for other laboratory members, especially to an anonymous crowd of people doing the analysis:

> Where it's like, we are used to doing it this way, just manually, and when I do it manually, I trust it. But human computation, well, the wisdom of crowd type thing is like okay, well, you trust it, but is it right?! And statistically, it [manual analysis] is less likely to be right than getting a whole bunch of people that maybe don't know all the details of things, but you take an average of all those people, creating something that is probably more accurate than what you would do […] the slow, painful way. [B]ut I think there is still a little bit of a mis… […] I wouldn't say a lack of trust but just more of a strong belief in their own abilities than handing things off to other people and then taking those results and […] taking those as a given. Definitely […] with programs that they are used to […] they can plug things in, and they get an answer from that, and they'll trust it if they've used it a few times. But the human computation stuff, they don't seem to think of it as another program. […] [A]nd there is like this bit of […] an inherent mistrust of "well it's just a bunch of people, how are they gonna do better than I could?!" And a lot of it is […], I think, in the end, statistically, they will do better […] than [researchers] could, but it's asking a large group of people to do a small repetitive task, and you don't have to invest as much time into that, but I think there is still this idea they get their results back and […]

> I need to double check this, I need to double-check and I can't just take it. (Oct. 14, 2021)

For many laboratory members, handing off the analysis step that had been part of the researchers' working routines to an anonymous crowd was difficult. At the beginning of the Stall Catchers project, the laboratory, together with the Human Computation Institute, had invested a lot of effort into validating the crowdsourced image analysis to the point where they were confident that the crowdsourced analysis was at least as statistically accurate as the researcher's results. Researcher Jada described this process as challenging because it involved a lot of extra and duplicative work:

> [I]n the beginning, it took at least […] three to four years to really kind of get something out that works. And this is a long time and … along this way, you did a lot of doubling of your work because I was counting […] as before just to make sure that this sort of matches up. And then you […], it's maybe kind of your baby, you're switching away to […] trust it. It's like, sometimes you have a person working with you; it's kind of the same. So, you do sanity checks, but then you give it away. But those times just have been sometimes months of work to […] confirm it on both hands. […] [But] I think it was still critical for the data and for me personally, mostly. (Oct. 27, 2021)

Eventually, Jada explained, researchers had to give away the task and start trusting the outsourced analysis. Still, researchers had trouble relying on the crowd's answers. There was a gap between the statistics and the distance between the platform and the crowd, as perceived by laboratory members. Jada further described the difficulty of no longer interacting with the data, of not being able to build the relation with the images they would usually build if they had annotated the images themselves. Looking at the images themselves is sometimes "just […] for [their] sanity" (Jada, Oct. 27, 2021). Jada explained, "I just would like to see these in the full stack" (Oct. 27, 2021). The reason they could not simply trust the result was in part because the preprocessing programs might still have difficulty identifying potential stalls in certain regions or edges (Jada, Oct. 27, 2021). However, Jada noted, "on a whole scale, you could say that doesn't matter, which is sort of true, but I still think it matters sometimes" (Oct. 27, 2021). Even if the difference was not statistically significant, it was important for the researcher's reason to analyze the images manually.

It was also a matter of routine. "[I]t's also the way I got trained myself to do it" (Jada, Oct. 27, 2021). Here, trust was built in and with routines and, as Pink and colleagues observed in their research on digital data anxiety and practices, "trust is invested in the routine, or a sense of trust is gained through the familiarity of the routine" (Pink, Lanzeni, and Horst 2018, 7).

Student researchers or those in their early scientific career "are all focused on future career and […] building up to something," explained researcher Charles (Oct. 14, 2021), who was further along in his scientific career. Therefore, there was another reason for them not to simply trust Stall Catchers and new computational tools: "So […] the ambition tends to make them a little bit distrustful of anything new, and I think that's both Stall Catchers and [new computational tools]. It takes a lot of work to get them to feel comfortable with it and to adopt it" (Charles, Oct. 14, 2021). Since their future scientific careers

depended on their performance in the early years of scientific research, it was crucial for them to maintain direct control over the research processes. This became more difficult when some key processes were delegated to other (unknown) people or to computational tools that were not established research tools.

Recalling the "Closing the Loop" section of the last chapter, I would like to discuss how this lack of trust in the results of Stall Catchers and the preference for manual annotation and engagement with the images unfolded in the researchers' everyday practices with Stall Catchers results. Once Stall Catchers had completed the dataset analysis, the crowdsourced annotation results of individual vessel videos as flowing or stalled could be downloaded by the researchers as a spreadsheet. Researchers could go back to the videos and "recheck" (Emily, Sept. 08, 2021) the crowd results with the links to the corresponding Stall Catchers video files in the spreadsheet. Some researchers even returned to the "raw" image files to verify the data after reviewing the Stall Catchers vessel video.

Such a validation step was not intended to be necessary in the original design of the Stall Catchers' collaboration. However, Michelucci explained in a conversation in October 2021 that the initial idea had been that once they had the crowd answers, they would no longer need expert answers. The crowd's answer was supposed to replace the expert answer, allowing researchers to take the results and move on to the next step in their research. In practice, however, researchers still went back to check all the stalls in the images analyzed by the crowd.[12] Consequently, a second protocol that they used during the time of my research was implemented to at least facilitate the verification of Stall Catchers results. This was the procedure discussed in the previous chapter, where the results of Stall Catchers were ordered from high crowd confidence of a vessel being stalled to low crowd confidence, so that researchers only had to review the first 200 videos rather than all of them (fieldnote, Oct. 18, 2021). As the researchers reviewed the results, they added their own vessel video annotations to the spreadsheet. These new annotations—not the participants' aggregated annotations—were then considered the final "ground truth" data labels. The crowd answers, or more specifically, the crowd's confidence levels, were, thus, considered as benchmarks and guiding indicators but not valid answers the researchers would continue to work with directly. Nonetheless, the participants' annotations greatly reduced the researchers' workload, who only had to review about 200 vessels compared to the full manual annotation of about 50,000 vessels in a dataset (fieldnote, Nov. 4, 2022).

This example shows how new trust-building procedures and practices were implemented to allow researchers to manually review the results and connect the abstracted vessel videos, the *dry* data, back to the *wet* data, i.e., a mouse model with or without a specific treatment to create confidence in the data. Trust in the results generated by the crowd did not simply exist but was being built by interacting with the data—in its different states from imaged mice to "raw" image stacks to Stall Catchers videos—and the resulting numbers and data references presented in spreadsheets. At the same time,

12 Similarly, in the example of Foldit, researchers reviewed all solutions submitted by participants by both manually looking at them and with computational analysis. Foldit researcher José explained that they did this "to try to see if I can find any errors just from looking at their models, but if they look good then we can test those in the lab" (Jan. 22, 2020).

trust in the CS collaboration was established in researcher–technology relations through infrastructuring, by working on the individual preprocessing steps, becoming familiar with the functionalities of new tools, and building a trustworthy data pipeline that would output the correct number of vessels (Chapter 6). Trust-building practices related to Stall Catchers were well underway but not yet established in 2021.

In the previous chapter, I showed how, during my second research visit in 2022, it was hoped that these would be established and, thus, the Stall Catchers loop closed once the new post-processing tool, which was supposed to further facilitate the researcher's experience with the pipeline and Stall Catchers, was in use (fieldnote Oct. 25, 2022). The hope was that once this new tool became part of the everyday research practice, all laboratory members would finally trust the whole project; building trust takes time (Endreß 2012, 91). The example of the data pipeline shows how trust in the collaboration with Stall Catchers "moves in and out of different idioms" (Corsín Jiménez 2011, 183), such as data quality, vessel count, and result-checking, and cannot be reduced to trust or lack of trust as a mere social relation.[13]

Another example of how previous trust-building practices had to be reconfigured with the introduction of Stall Catchers is the "blinding" procedure, which had to be translated from its original manual form into the virtual game environment. As described above, researchers would typically analyze image stacks manually without knowing the treatment or mouse model on which the data were based. However, despite efforts to de-identify the images, it was still possible for some researchers to identify the underlying mouse model:

> [T]ypically, I can recognize an Alzheimer [mouse] relatively easy. […] [G]enerally, that doesn't mean I know the treatment. I try to not care at this stage about it […] or [another researcher] is doing it and [this other researcher] definitely doesn't know what mouse is what. But for me […] because I have done many other analyses with those mice before[ ], I typically know what it is. (Jada, Oct. 27, 2021)

The introduction of Stall Catchers, therefore, presented an opportunity to even improve anonymized analysis. At the same time, however, translating the anonymization process into the virtual game platform posed several challenges. For one, the crowdsourced data analysis approach was not an established research method with defined procedures. Therefore, the laboratory had to work with the Human Computation Institute to develop a new process to guarantee that participants would not know what research data they were looking at. It was essential that the original data, i.e., a mouse with or without Alzheimer's disease, be alienated. Nishimura explained:

13 Interestingly, when I returned to the laboratory in Fall 2022 and discussed my observations regarding trust with the biomedical researchers, some mentioned that they still preferred to sometimes annotate research data manually instead of sending it through Stall Catchers. When reflecting further about this preference, one of the laboratory members explained that there was no longer any reason for this, since the problems they used to have with ensuring the data quality with the Stall Catchers data pipeline had been resolved (see Chapter 6). However, the gap had not yet been closed, the new process of preparing data and sending it through the crowdsourced analysis had not yet become an established routine (fieldnote, Oct. 21, 2022).

> [I]t took a few iterations to come up with […] this blinding, […] that's the gold standard for data analysis, and sometimes we don't even quite do it in the lab as well as I would like. […] So, it kind of takes a bit of […] a leap of faith that you really have to do this right, but I do think it's important. (Dec. 7, 2021)

To get it right, in terms of scientific data quality standards, it was necessary to "sacrifice" user or play experience, as I discussed in Chapter 5. While they first thought about including a progress bar for individual mice, they later refrained from implementing it to ensure that "the whole dataset is blinded" (Nishimura, Dec. 7, 2021).

As the analysis of the intraverting researcher–technology relations showed, however, incorporating crowdsourced analysis of their data also required the development of a data pipeline, which was incomplete with the introduction of Stall Catchers and an ongoing endeavor. While I have already analyzed the process with a focus on the changing human–technology relations in the data pipeline, I here would like to return to this example with a focus on trust. In 2021, according to Schaffer, the laboratory had come a long way regarding the Stall Catchers project and now "already [had] a good confidence in Stall Catchers" (fieldnote Aug. 17, 2021). Compared to the early days of Stall Catchers, they now had "good buy-in" (Schaffer, Dec. 7, 2021). During our interview in December 2021, he noted that in the beginning,

> we had a period where people, I think, had not as good a buy-in on Stall Catchers because they saw it as a lot of upfront kind of work that they had to do. And then … there was lingering uncertainties about the data quality. But I think with people being actively involved in fixing those problems, I think that's how you get by it. […] And so now people can work with it and understand the limitations, understand the capabilities and not feel […] so uncertain to or untethered, I guess, in their use of that capability. (Schaffer, Dec. 7, 2021)

Active involvement in the process, the data pipeline, and the individual tools created confidence and trust. Therefore, for example, laboratory members were required to complete training on laser alignment, which included learning how to set up the lasers for the microscopes. Researchers usually did not have to set up the laser parcourse from scratch for imaging, as it could be reused from previous imaging sessions and did not need to be changed for each individual one. Nevertheless, they should know what is going on behind the technologies supporting their work to be able to adjust them if necessary and, as Schaffer explained, to not "overtrust" the infrastructures (fieldnote Oct. 20, 2021).

In summary, prior to the introduction of Stall Catchers, building and maintaining trust in the "right" analysis of the research data and its results had already been characterized by sociomaterial practices, such as going through the image stacks, focusing one's eyes on the vessels in the raw image data, and, thereby, interacting with the data in a way that one could rely on one's own abilities and practice. Building trust was more straightforward here than with Stall Catchers, also because the data was considered "raw," or, at least, not manipulated by computational algorithms (although it had already been translated from mouse brains into digital images). Knowing that they were looking at the "original" data, there was no room for doubt about what they were seeing. The data rep-

resentation displayed the "real" data, and no bug in any algorithms could have introduced distorted images or missed some vessels. Additionally, extra measures, such as the independent analysis of the same image stacks by multiple researchers and the anonymization of the research data, ensured "scientific confidence" (Schaffer, Dec. 7, 2021). With the introduction of Stall Catchers, not only did new collaborations with the Human Computation Institute and an anonymous crowd of people disrupt established practices and require new ones, but the introduction of new data pipeline steps for Stall Catchers, in particular, introduced new potential sources of error and, therefore, uncertainty suddenly multiplied.

Let us now return to the opening quote of this chapter after the analysis of how trust was built and maintained in scientific knowledge production with Stall Catchers at the biomedical engineering laboratory, namely, that they had "to trust the whole citizen science stuff" (fieldnote Jul. 27, 2021). It becomes clear that what was described in this rather plausible statement is not a mere social relation that is chosen instead of other approaches but unfolds in sociomaterial practices along intraverting human–technology relations. These practices, however, are to be located on the laboratory's side of the Stall Catchers project. Regarding Stall Catchers as the game, the software, and the HC system itself, the biomedical researchers were "confident" (fieldnote Aug. 17, 2021) and relied on the Human Computation Institute. Here at the institute, I observed other trust-building practices, which I discuss in the following section.

Building Trust With Human Computation and Algorithmic Evaluation

When I asked Paul, who had been part of the Stall Catchers team since its early days, how HC-based CS games like Stall Catchers (could) change science structurally, he explained that these projects "contribute—I hope they're contributing—to changing […] the mentality, the way people think about science but not just normal people but the scientists themselves" (Oct. 14, 2020). Questions of trust in research with CS in general and how CS could change established structures and hierarchies within science have been discussed in scientific and public discourse (e.g., Haarmann 2013; Kosmala et al. 2016; Bedessem, Gawrońska-Nowak, and Lis 2021), and there are various proposals for frameworks to build trust in CS. For example, computer scientists Abdulmonem Alabri and Jane Hunter discuss a technological framework that uses trust models and filtering services to improve the reliability of and trust in CS (Alabri and Hunter 2010; *cf.* Hunter, Alabri, and van Ingen 2013). But despite the popularity of CS today and the support by major funding programs, such as the European Union's Horizon Europe 2021–2027 (European Commission 2021), recurring concerns remain, including those related to the evaluation of results. The authors of the Science Academies G7 Summit report on digital CS express concern about the risks associated with CS, "especially around the evaluation of results stemming from CBPR [Community-Based Participatory Research] and BTWR [Beyond The Walls Research]. These results are often disseminated through diverse channels outside the traditional peer-review system" (Gaffield et al. 2019, 1). Human Computation Institute team member Paul, who also observed such concerns, traced them back to an issue of mistrust on the part of some professional scientists. After his initial positive outlook on the potential of CS (quoted above), he clarified that today, the potential of CS has

not been fully explored, in part because of this mistrust: "I think it's […] an intimidating thing for most scientists to open up to the public […] and involve the public […] and they don't even understand what the public knows and can do and stuff like that. They, *they* mistrust the public" (Paul, Oct. 14, 2020). Even if CS projects were to follow standard scientific validation techniques, the problem of professional scientists not being familiar with the approach and methods of CS and the involvement of nonprofessionally trained scientists in research projects, in general, would remain. This is where projects like Stall Catchers, especially as they have spread in recent years, could play an important role in building trust, Paul argued (Oct. 14, 2020). With Stall Catchers, they "have actual evidence to support […] this method […] and to […] demonstrate data quality and all the other stuff that scientists […] worry about," continued Paul (Oct. 14, 2020). Developer Kate agreed with Paul in our November 2021 interview, clarifying that "we're proving that […] you can do things differently" (Nov. 19, 2020). Although more and more institutions, libraries, political bodies, and universities were beginning to adopt a more open attitude toward CS, Paul did not expect a rapid change in thinking and academic scientific practice because academic institutions "change over [the course of] hundreds of years […] but not several years" (Oct. 14, 2020). The COVID-19 pandemic has also played a role in accelerating the popularization and visibility of CS since early 2020, driven particularly by successful projects such as Foldit. "[T]heir visibility and their success, I think, is […] helping scientists to understand what it is to open up and why do it, so … I'm kind of … finding myself sounding too optimistic [laughs], but I do wanna believe it's making that sort of a difference" (Paul, Oct. 14, 2020). Here, the Stall Catchers team argued that HC itself was "key" to gaining the trust of professional scientists in CS:

> I think that […] this method [of HC] is the key to ensuring these things are done right. Cause you can, I mean, anybody can create a citizen science project […] and just get people out there and do something. But unless you understand how to get the maximum value of individual contributions, then you're mostly just wasting your time and such a project I think risks to just become another science outreach project. […] I'm sure there might be other methods, but I'm a bit biased, so I think human computation is sort of key in these types of projects, and […] we will not open science unless scientists can trust the way it's done, which is why a … solid scientific foundation like human computation methods must be behind it, I think. (Paul, Oct. 14, 2020)

HC was understood to provide the computational and statistical methods to bridge the gap between scientific knowledge production, which depends heavily on measurability and calculability in the natural sciences, and CS, as the engagement of the broader public without predefined knowledge. It, thus, linked people to numbers. Here, building trust with HC as a computational method carried the notion of legitimizing CS as a scientific approach.

But how did HC actually contribute to building trust within these projects? If biomedical researchers had to invest in infrastructuring and new working practices to reestablish trust in the results of analysis, what mechanisms, relations, and practices played a role within the human-in-the-loop system itself?

In a conversation with developer Samuel, he explained that developing Stall Catchers required a careful approach: "Because since we're working with the crowd, we're like asking everyone to connect to our system and [...] all this then can be hacked and manipulated. So, we have to be careful" (Sept. 2, 2021). As an Internet platform that was meant to be accessible from all over the world and to everyone, Stall Catchers, like any other freely accessible website, was vulnerable to both targeted and random attacks. However, as Samuel explained, caution was especially important in the case of Stall Catchers for another reason: it worked with a crowd of people the team did not know. Stall Catchers developers also described this as a specificity of developing HC systems compared to other software projects they had worked on. According to Samuel, in other software projects, such as the development of online rating or booking platforms, it is often easier to identify the target audience and anticipate their needs on the platform. Moreover, since cheating is an inherent part of gameplay (see Chapter 5), the Stall Catchers source code included protection mechanisms to prevent cheating. Cheating, however, could not only break the game's functionalities and cause unfair play but could also lead to poor data quality.[14] This is probably the most significant aspect of the role of trust in HC-based CS projects and how trust is built and maintained at the source code level because poor data quality could harm the overall research. Since Stall Catchers contributes to Alzheimer's disease research, the team's top priority was to ensure that the quality of the analysis results met the requirements of scientific research as defined by the laboratory. To achieve these goals, the Human Computation Institute developed a customized "wisdom-of-the-crowds" method to calculate "final" Stall Catchers answers for a given data point. This method, as described previously, combined the responses of several Stall Catchers participants for that data point and applied a weighting mechanism to these responses according to the participants' "sensitivity" scores. This means that building trust in the collective answers and meeting the required scientific data quality requires customized algorithmic mechanisms. Before including the answers of new participants in the collective crowd answers, their skill level was assessed in the pregame tutorial. Once the tutorial was completed, participants could annotate "real" research videos for which there were no expert answers, and their annotations were incorporated into the calculated crowd answers according to their skill level. This skill level was not fixed, however, but constantly recalculated based on the participant's answer to "calibration movies," which were regularly sprinkled between research videos and to which the correct answer was known. These calibration videos were intended to be and usually were indistinguishable from the research videos. However, experienced participants and so-called "supercatchers" were sometimes able to tell the difference between calibration and research videos. Moreover, the frequency with which these "calibration movies" appeared was not random or the same for all participants but dependent on the individual's skill

14 Cheating in HC-based CS games can be categorized as either harmful or valuable for achieving the scientific purpose, as illustrated by the example of cheating discussed in Chapter 5, where a shortcut in the game discovered by a participant allowed the acceleration of data analysis. However, not knowing the intentions of individual participants and how cheating might influence data quality, cheating-prevention mechanisms had to be put in place. This distinction was also discussed on the Foldit forum in 2014 (v_mulligan 2014).

level. This shows how different participants and their contributions in Stall Catchers were not treated equally by the system but, on the contrary, how the system was tuned to individual participants. Different levels of trust were introduced algorithmically by continuously measuring the skill level and weighting answers accordingly.

Figure 11: Excerpt of the Stall Catchers source code with the saveNextMovie function

```
/*
 * Save next movie (for criminal clients who want skip hard known movies by
 * turning off browser
 * @param int $movieId
 * @param int/false $nextMovieIdToPreload // movie to preload in view
 * @param int $userId
 * @return bool
 */

public function saveNextMovie($movieId, $nextMovieIdToPreload, $userId)
{
    if (empty($userId)) {
        return false;
    }
    if ($nextMovieIdToPreload === false) {
        $updateArray = ['next_movie_id' => (int)$movieId];
    } else {
        $updateArray = ['next_movie_id'         => (int)$movieId,
                        'preload_next_movie_id' => (int)$nextMovieIdToPreload
        ];
    }
    $this->model->where('user_id', $userId)
                ->update($updateArray);
    return true;
}
```

Source: ©Human Computation Institute n.d.

In addition to the encoded trust-building practices to ensure scientific data quality, there were other algorithmic mechanisms in the game's source code to prevent cheating and, hence, harm to data quality, which I will discuss here with the example of the "saveNextMovie" function (see Figure 11). The purpose of this function is described in the comment above the function as "Save next movie (for criminal clients who want skip hard known videos by turning off browser." The function prevents users from switching to another vessel video without answering the current one. The algorithm, thus, ensured that all participants and their corresponding skill levels were evaluated on the same video data. The code here included the design assumption that some participants—specifically "criminal clients"—would actively try to circumvent the rules. Instead of assuming that participants would answer all videos as presented, the source code included this function to ensure that skipping videos was impossible. Furthermore, this fragment shows that it was assumed that "criminal" participants would strive for points and winning the game rather than focusing on contributing to research by skipping those videos that might be more difficult to answer correctly. Hence, decision power over the selection of the next video clearly rested with the algorithm, not with the participant.

The question of how to establish trust in HC systems is not answered once and for all but evolves alongside the continued development of participant–technology relations in HC systems. Intraversions in participant–AI relations, as with the introduction of AI bots in Stall Catchers, require new practices for building trust, as explored in the experimental study on human–AI partnerships in 2020 (Inkpen et al. 2023). This experiment, in which an AI agent was introduced into the Stall Catchers task to assist human participants with suggestions about whether a vessel video was flowing or stalled, explored the performance of hybrid human–AI teams depending on the skill level of the human participant and the sensitivity and bias of the AI agent. Michelucci explained the goal of this experiment in one of our conversations:

> [W]e […] explored an actual collaboration […] So I have this agent that's assigned to me as my partner. That AI agent has some level of sensitivity and some kind of bias with respect to answering about stalled and flowing. And the AI is giving me suggestions. And now, I have my own ideas about what's flowing and stalled and based on […] seeing what the AI says, making my own decisions, seeing the outcome, I start to develop a trust maybe that the AI in certain cases knows better than I do, but in some cases, I know better than the AI. And if I could develop that predictive model about when the AI is going to be right or wrong and integrate that with my own responses and maybe collectively as a dyad, we can perform better than individually. And […] our early result is that when the machine bias is opposite of the human bias, then we actually see better performance with the AI and the human working together than separately. So that was exciting. (Jan. 21, 2021)

While this experiment took place in a sandbox (a special environment designed to test an isolated feature or experimenting without all game functionalities and with no direct impact on the real game or Alzheimer's disease research), AI bots were later introduced into the actual Stall Catchers game, as I discussed in the previous chapter. Even before the introduction of AI bots, the idea of human participants and AI bots playing Stall Catchers side by side had already raised many questions and concerns. A few months before the AI bot GAIA was introduced to Stall Catchers, Paul described his expectations for the introduction of the bots:

> [There will be] some people who will be interested and [who] more or less understand what's happening and even be interested in competing with AI or whatever, and then there will be people who will be confused and feel replaced and not sure, mistrust the AI, and maybe the AI is now messing up the data, and their work will also be wasted. (Oct. 14, 2020)

Human–AI relations in the experimental bot studies unfolded in productive team spirit and competitive relations. How these relations will evolve in the future, if and when AI bots become a permanent part of Stall Catchers, remains to be seen. But this example shows that, much like human–technology relations, trust is not built once and for all but changes along intraversions. Finally, although the understanding of trust as a sociomaterial practice describes a shift away from the conceptualization of trust as a purely cognitive phenomenon, it does not exclude trust as a social relationship as defined by

Luhmann (1988; 2014) and others (e.g., Hardin 2006; Weingardt 2011). As a final example of trust from the perspective of the Human Computation Institute, I would like to show the importance of building trustworthy social relations for the development of new HC-based CS projects for researchers and developers. During my fieldwork in Ithaca, I accompanied Michelucci on a two-day trip to collaborative researchers in Washington D.C.[15] at the end of September 2021. The main goal of this trip was to "build relationships" (fieldnote Sept. 30, 2021) with researchers as partners for a new HC project. We left Ithaca in the late morning of September 29, arriving just in time for dinner with the researchers. To my surprise, the envisioned project was not discussed that evening. The next day, after breakfast, Michelucci and I reflected on the first impressions of the visit and discussed the importance of building social relations. Michelucci mentioned that he often had to first explain the Human Computation Institute's projects, such as Stall Catchers, and the organization of the Human Computation Institute because "the HCI itself is an unconventional institution" (fieldnote Sept. 30, 2021). Therefore, meeting in person and spending time together to get to know each other and build trustful relationships before developing a new HC-based CS project was important. While they had started working together and developing ideas for new projects more than a year ago, it had not really taken off before the September 2021 visit. The face-to-face meeting made it easier to discuss mutual expectations and clarify the approach, which could build upon the Human Computation Institute's previous experience in building Stall Catchers but for which there were no common procedures. A trusting relationship could be established through a personal meeting.

Trust the System, Trust Yourself

If we now turn to the participant's point of view, trust and mistrust unfold yet again differently. Trust came up in several conversations with participants when I asked them how important it was to them to learn about the scientific developments behind Stall Catchers. Here, trust in the Human Computation Institute and the researchers conducting the Alzheimer's disease research played an important role. At the same time, trust, or more accurately, mistrust in one's own abilities and in getting the answers right, was either explicitly or implicitly raised by some participants as one of their concerns. In this section, I focus on these two dimensions of trust.

Using the Human Computation Institute's blog, the Stall Catchers team provided insights into what was going on behind the UI at the institute, announced new features and special events, and provided updates on the research behind the game. Whenever a new dataset was uploaded to Stall Catchers for analysis, a blog post was published explaining the purpose of that particular dataset. Institute member Egle (seplute) explained in a July 22, 2017 post that "[t]he new dataset is focused on the effects of a **high fat diet on stalls in the brain in Alzheimer's disease**" (2017, emphasis i.o.). This brief statement was followed by an explanation of the research question or aim and its impact on Alzheimer's disease research: "We are seeking to understand the cellular mechanisms linking cardiovascular risk factors to Alzheimer's. Analyzing this dataset will be a big push towards un-

15 The destination is anonymized.

derstanding this long-debated link!" (Egle [seplute] 2017). For participants who wanted to learn more about the dataset, a more detailed scientific explanation was provided, citing research collaborator Schaffer. The Stall Catchers team aimed to make the research behind Stall Catchers more comprehensible and accessible with these blog posts, which were also shared and linked in the in-game chat.

Interestingly, the answers varied widely when I asked participants about the importance of learning about these developments and understanding the science behind Stall Catchers. It was important for many participants to learn about the scientific developments because they "like to know how playing the game is actually helping progress the research (Ebby, May 08, 2020). Or because "[t]hat's a reason […] for what you do. […] That's certainly something you can learn from" (Kamon, May 15, 2020). Akin even doubted that they "would engage in it at all if it weren't for that" (May 11, 2020). Reading about the progress created a sense of accomplishment and ownership: "Well, it's important from the standpoint that if I see the science of this making progress forward, I can feel a little piece of ownership that, hey, I think I've helped with making that progress forward" (Daan, May 26, 2020).

Other participants, however, expressed their interest in learning about the datasets they were analyzing and the scientific developments as a form of feedback that was not particularly important to them: "for me [it] is really not that important because I am confident that it will get sent there somewhere," Ellen explained (May 19, 2020). Caitlin said,

> it makes it more interesting if you know how the data is being used and what the scientists are learning from it. So although it wasn't my primary motivator early on, I just trusted that what we were doing was of use or we wouldn't be doing it, but it's good to get the feedback like with this NOX-inhibitor business, […] that's very positive feedback to be getting that […] we might be making some progress here and in some small way what we're doing may help advance progress towards a cure perhaps even for Alzheimer's. So, yeah, that's very good to have that kind of feedback. (Caitlin, May 5, 2020)

It was not particularly important to these participants to be kept informed of all scientific steps, or it was not the primary motivator for learning about Alzheimer's disease research because they were "confident" or "trusted" that their contributions would be meaningful and used for scientific purposes. The terms "trust" and "confidence" were used here to describe a similar relation. Although "trust" and "confidence" are commonly distinguished in sociological theories (among others, Luhmann 1988; 2014; Hardin 2006), the purpose of this work is not to apply a fixed definition of these terms to the empirical material but to better understand how trust unfolds. With this in mind, it is possible to see how trust navigates through various expressions in everyday life.

The distribution of knowledge in Stall Catchers and the role allocations of different actors can explain the trust relation. Longtime participant and frequent player John explained that

> you have to be able to trust the people […]. Hopefully, the people at the top would be willing to say "Hey, we've kind of reached the end of where we wanna go here;

> maybe we gonna have to take some time moving to another area of research, or maybe it's time to pull this back"… or […] if the people at the top are honest and if you believe they are—cause it's hard for maybe the average person to verify—[one would have to] read a bunch of articles […], you put your head down…and it's not! I don't mean to make it sound like a struggle, but you dive in, and if you want to do it, then you do it. And if you do one film, that's good! If you do one million films, that's good! But we're all trying to go towards the end of hopefully—whether we know it or not—figuring this thing out. Even […] if the research shows something contrary to what the scientists had thought, it was going to show and it's a dead end, that's good too because we know where not to look now and hopefully we spread that around saying you know "we did this and it is peer reviewed, and there is no point really going down this place anymore." (John, May 7, 2020)

While participants could read about the science behind Stall Catchers, they did not need to know the details about Alzheimer's disease research to contribute to the project. They had one specific task on the Stall Catchers platform: to analyze the data presented to them. While the Human Computation Institute took care of the algorithmic evaluation of participants' input and ensured that the required scientific data quality was achieved, the researchers in the Schaffer–Nishimura Lab directed the scientific approach, including decisions on which the research questions to investigate, and how to proceed.[16] Here, trust in the researchers filled the gap in scientific knowledge about Alzheimer's disease research and allowed participants to perceive their contribution as meaningful.

Despite trust in the scientists and the Human Computation Institute team to develop and maintain the platform and game in such a way that the participants' contributions were perceived as purposeful, I observed another dimension of trust in the participants' own abilities to analyze data on the Stall Catchers platform. Some participants expressed forms of mistrust in themselves or, more precisely, in their own ability to make meaningful contributions to Stall Catchers and Alzheimer's disease research. Some participants explicitly described their fears of answering research videos incorrectly in our conversations. Elle, for example, who began contributing to Stall Catchers in 2015, felt that she did not have the time to participate in Stall Catchers as much as she would like because her "level of emotional investment and commitment isn't always matched by [her] actual actions" (May 13, 2020). At one point in our conversation, she expressed fear of being presented with a research video that had not yet been annotated by many other participants (the UI includes an information box below the video frame indicating how many participants have answered the video presented and how): "when you start moving from the ones that are being done by multiple people into ones that you're the first person or the second person to check and you don't actually know whether other people agree with you or not, it's a little bit more scary" (Elle, May 13, 2020.).

Noemi, a participant who had not contributed to Stall Catchers for a while before our interview but who had gone back to Stall Catchers the night before our meeting, explained as she reflected on her Stall Catchers experience: "I was doing some last night, and I was like oooh now I have to get back into what I'm looking for … cause I'm making

16 As described by Michelucci, in a few rare instances, though, the researchers took up suggestions from participants for specific studies (fieldnote Nov. 9, 2023).

mistakes, and I'm feeling bad, I'm making mistakes, that's wrong! But usually when I have long periods of time is when I will go and play" (Noemi, May 14, 2020).

Stall Catchers participants contributing to my research agreed that becoming "good" at Stall Catchers required practice. Noemi further explained that she would have to get used to playing Stall Catchers again and attributed her mistakes to her long absence from the project. It was very important for her to contribute in a meaningful way, and, therefore, she was afraid of getting videos wrong: "[I]t is challenging! It is challenging and sometimes […] with this game I'm like 'oh no I don't wanna get it wrong'" (Noemi, May 14, 2020).

The fear of contributing "bad data" can be connected to the participants' understanding of how the HC system works and their practice, i.e., how often and how much they contributed. Participant Olav explained how participation helped build trust in one's annotations: "the longer I think you do it, […] you […] can then trust yourself after you've made a decision […] that's the right decision because you've done it enough to know […] what […] clear, flowing […] things should look like" (May 21, 2020). By regularly engaging with vessel videos, participants learned to distinguish between flowing and stalled vessels and developed an understanding of the system's functionalities. According to the Stall Catchers team, the implemented security measures and the "wisdom-of-the-crowds" methods made it virtually impossible for incorrect answers to harm the overall analysis. Most of the participants who were aware of this did not fear inputting incorrect data, although some also described missing stalls as a bad experience (see Chapter 5). In an interview with the Human Computation Institute published on the institute's blog, "supercatcher" Carol aka Mema described her "misses" as one of the things she disliked most about Stall Catchers because "I feel like I let myself and others down" (Carol aka Mema 2019).

Practicing and engaging with the system regularly, as well as reading up on the features in the FAQs or blog posts explaining the design of Stall Catchers and, thus, understanding how Stall Catchers as an HC system and the "wisdom-of-the-crowds" method work can, therefore, be interpreted as trust-building practices (*cf.* Pink, Lanzeni, and Horst 2018). Participants do not have to worry about their skill level in catching stalls, as participant Maya summarized: "[I]t doesn't matter if you're right or wrong because it's all statistics and […] it's okay to be the wrong one, and it's okay to be the right one. And then depending on if you're right or wrong however a number, many of times, it starts to change what they give you" (May 13, 2020). By contrast, lacking this experience and the understanding of the game mechanics and crowd answer computation led to not trusting one's own analytical skills to contribute meaningfully to Stall Catchers.

The Question of Trust and Proprietary Software

Before turning to concluding thoughts on trust in and with HC-based CS, I would like to address a final aspect related to the participants' ability to understand the system and, hence, to verify that their contributions are meaningful. This aspect does not only apply to the example of Stall Catchers but was also discussed in the example of Foldit.

Both projects are based on proprietary software and the teams justify this in part because it is a game and people might try to modify the code in ways that are beneficial to the game but harmful to the science behind the projects.[17] Foldit team member bkoep explained in a forum discussion that "[p]art of the concern with open-sourcing the Foldit code is the potential for abuse—there are a lot of trivial ways that Foldit could be made more 'fun' as a game, but that would also undermine the scientific validity of Foldit players' work" (Bkoep 2018). In Foldit, proprietary software was further justified by the fact that Foldit was built on the code base of the protein-modeling software Rosetta. The Rosetta code was distributed under a Rosetta license, which was not open-source (but free for noncommercial use) and required registration with RosettaCommons (Bkoep 2018). Following Giddens, the "prime condition of requirements for trust is […] lack of full information" (1990, 33). Participants in HC-based CS games that rely on proprietary code, such as Foldit and Stall Catchers, cannot verify scientific correctness even if they wanted to and had the technical knowledge to do so. Stall Catchers participants, for example, cannot verify that the statistical approach to combining individual answers is indeed correct. Instead, they must rely on the team of the Human Computation Institute and believe the system's explanations, which are simplified to make them accessible but, therefore, also incomplete. This is not to say that there is a reason to distrust the developers and researchers but to demonstrate the existence of this gap.

While the question of open-sourcing Stall Catchers had not been raised publicly on the Stall Catchers forum at the time of my research, in the case of Foldit, requests to open-source Foldit's source code had been recurring since the project's early days. Now and then, a new participant would (unknowingly) reopen the year-long discussion, prompting some other participants to leave comments such as "here we go again" (B_2, 2011), expressing their fatigue with the discussion. Nonetheless, these requests kept coming back and, most of the time, did not proceed satisfactorily for the requester. It is not surprising that such requests are more common at Foldit, as it attracts many programmers and participants interested in computer science. Since the code base of both projects is not open-source, it becomes impossible for participants or others interested in them to build trust as sociomaterial practice.

Distributed Trust

HC-based CS assemblages are multiplicities resulting from continuous reterritorialization processes in which different interests and human–technology relations are aligned, together forming the assemblages. Continuity is necessary due to deterritorializing processes that simultaneously act upon and carry away the assemblage. Examples of the latter are divergent interests, different logics at play (as discussed in Chapter 5), or material breakdowns. Furthermore, as Deleuze and Guattari write, an assemblage "necessarily changes in nature as it expands its connections" (2013, 7). Due to the need for

17 Interestingly, the ARTigo source code, by contrast, is open-source and published under the GNU General Public License on Github (Institute of Art History (Ludwig Maximilian University of Munich), n.d.).

HC-based CS systems to remain at the edge of AI, human–technology relations keep intraverting, and with them, the systems or assemblages themselves are changing. These intraversions and changes in the overall system require trust, an example of a reterritorialization process, to be continuously rebuilt. Trust, as I have aimed to show in this chapter, thereby, moves along various sociomaterial relations and practices. In the examples discussed, trust cannot be grasped by analyzing human social relations alone, even though they play an important role, as the example of the trust-building trip I accompanied demonstrated. Trust is built via human–technology relations. Thus, even if the explicit goal of the team members was to trust, for example, the participants contributing to Stall Catchers, this was achieved through the mediation of technologies and sociomaterial practices. It is, therefore, important not to overlook trust as a sociomaterial practice, which was the focus of this chapter.

The introduction of Stall Catchers in the biomedical engineering laboratory required the development of a data pipeline to prepare research data for subsequent analysis by Stall Catchers participants, which partly disrupted established research practices and work routines and even led to mistrust in the results. Trust-building practices, such as anonymized research data analysis, had to be translated into new practices. Trust emerged here through work on infrastructures and engagement with materialities.

Additionally, trust played a role in legitimizing CS as a not-yet-established approach to scientific knowledge production. Here, the HC system mediated trust in crowd answers via computational algorithms by preventing cheating and ensuring the required scientific data quality. At the same time, trustworthy relations with collaborators need to be built first, e.g., through in-person meetings and time spent together, to develop new HC-based CS projects.

Finally, a lower level of familiarity and understanding of Stall Catchers on the part of the participant tends to increase the importance of trust in the researchers and the team of the Human Computation Institute in order for the participants to perceive their contribution as meaningful. Similarly, regular participation and engagement with the system build trust in the participant's ability to catch stalls. Participants who were less familiar with the "wisdom-of-the-crowds" method correspondingly expressed mistrust in their competence to contribute to research.

Building and maintaining trust is essential for Stall Catchers to come into being in the interplay of the different human and nonhuman actors involved. As I have shown, the relations between, for example, infrastructure, software tools, and researchers, or participants and algorithms are just as important as those between human collaborators. Trust, therefore, unfolds in (sociomaterial) practices. Following social anthropologist Martin Holbraad's Truth in Motion: the Recursive Anthropology of Cuban Divination (2012), in which he argues for conceptualizing truth in "motile terms, as an event of collision—a meeting—between previously unrelated strands of meaning" (2012, xxiii), I suggest thinking about trust in motile terms. Through engaging with meanings of trust in the field, trust reveals itself as not just a static phenomenon but one that is constantly becoming through these sociomaterial relations and practices. It, thereby, depends on the very sociotechnical situation and "transforms itself in various stages or steps" (Harper 2014b, 307) in HC-based CS assemblages. "[T]rust is transitive; certainly a composite" (Harper 2014b, 307). In this way, trust unfolds as something distributed

across HC-based CS projects, emerging within human–technology relations and along their intraversions, as much as it contributes to forming sociotechnical assemblages and holding them together. Finally, trust needs to be continuously maintained and rebuilt through the ongoing evolution of these HC assemblages.