2. Seeing through Machines: Algorithms and the Pursuit of Mechanical Objectivity in Embryo Imaging

MANUELA PERROTTA

Introduction

Biomedical imaging has been extensively explored by science and technology studies (STS) scholars (Lynch and Woolgar 1990; Coopmans et al. 2014). A rich literature has shown the situated complexity of producing images in biomedical research laboratories (Prasad 2005, 2007; Joyce 2006; Alac 2008; Burri 2008), including the importance of local context of production in constituting their meanings. Among these studies, a growing body of work (for a review see Perrotta 2013) has focused on the invisible work done by researchers to produce authoritative images (Joyce 2005; de Rijcke and Beaulieu 2014) and how images become visual evidence (Beaulieu 2002; Roepstorff 2007; Beaulieu et al. 2012). How images become "true" representation of the world is not a new question in science and has been explored in different historical contexts (see Daston and Galison 2007). However, the development of digital imaging and image-processing algorithms has introduced new concerns regarding their use in various settings, including medical practices (Reich 2012; Torenholt and Langstrup 2021).

In this chapter, I contribute to the body of STS literature exploring technoscientific imageries, one of Marina Maestrutti's main research interests (see Maestrutti 2008, 2011), by analysing the introduction of algorithmic technologies in the production and use of embryo images in fertility treatment. The case of embryo imaging is a prime exam-

ple of how the advancement of IVF imaging technologies is reconfiguring not only how embryos are visualised, but more generally the processes of knowledge production in embryology (Geampana and Perrotta 2021). In the last few years, conventional incubators have been replaced with time-lapse incubation and imaging systems (henceforth TLIISs). Time-lapse is an old imaging technique, where a camera is set to record a series of images at regular intervals. Before the introduction of TLIISs, an embryologist had to physically remove the developing embryos from the incubator for a brief period and check them each day under a microscope. The main advantage offered by TLIISs is that they allow embryologists to take thousands of images of the embryos (one every five to 20 minutes) without disturbing them, as they do not have to be removed from the incubator. In this way, TLIISs grant a continuous view of each embryo as it develops, rather than just viewing them once a day. In addition, TLIISs can be used together with embryo assessment and selection (EAS) algorithms with the aim of selecting or assisting embryologists in the selection of the embryos (I will discuss the difference between these approaches below) that most likely will develop into a baby. Due to the lack of evidence supporting their effectiveness in increasing the chances for couples of having a baby, TLIISs have been highly criticised both in the medical and in the social science literature (see Perrotta and Geampana 2020). However, little attention has been paid to how the use of algorithms reconfigures the production and use of embryo images.

The algorithms used to assess and select embryos accord with Gillespie's (2014: 167) definition of «encoded procedures for transforming input data into a desired output, based on specified calculations». As I explore in this chapter, EAS algorithms are computational and run automatically via proprietary software, although with different levels of opacity (Burrell 2016), playing an increasingly important role in defining what information is most relevant not only for embryo selection but for embryology knowledge more broadly. I explored elsewhere (Geampana and Perrotta 2021) the reconfiguration of knowledge production processes due to local algorithmic practices resulting from the increased use of TLIISs in fertility treatment. In this chapter, I focus on analysing the reconfiguration of professional expertise involved in «seeing through machines», by exploring how professional communities of embryologists have engaged with the ongoing pressure to overcome the subjectivity of visual judgment in embryo selection. In doing so, I aim to contribute both to the literature on technoscientific imaging and to the growing body of work exploring how algorithms are contributing to the transformations of expert fields in multiple ways (Pasquale 2015; Ananny 2016; Zarsky 2016).

Tensions between different forms of objectivity and expert judgment

As other scholars working on the reconfiguration of professional expertise emerging with the development of algorithms have pointed out (Christin 2016), the discussion on how

machines are mobilised in expert fields to support the production of certain, true and objective knowledge is not new (see Ziewitz 2016).

Science historians Daston and Galison (2007) have showed how the meaning of the concept of objectivity has fluctuated significantly over the last three centuries. Their study of the production of scientific atlases since the eighteenth century illustrates how different forms of objectivity in the visualisation of natural phenomenon were introduced and institutionalised over the years. The authors define as "truth-to-nature" the form of objectivity connected to realism that dominated for most of the eighteenth and nineteenth centuries. At the time, scientists' attempt to capture "nature" was achieved through depictions that were deemed able to represent the ideal of a certain natural phenomenon (for instance an animal or a plant). In practical terms, this meant that several specimens were often mixed to capture this objective ideal. For instance, the ideal representation of a certain plant was obtained through a collage of several parts of the plants that were considered ideal, but actually belonged to different specimens. This practice, however, was questioned over the course of the nineteenth century due to the active human intervention required and thus the introduction of human subjectivity and individual idiosyncrasies. These doubts pushed toward the use of new technological developments deemed able to minimise human intervention and produce "unmediated" representations of the real world. Daston and Galison define this new paradigm as "mechanical objectivity" and link it the widespread belief in the superior objectivity of machines such as daguerreotypes, photographs, and X-rays. In their analysis, mechanical objectivity represents an epistemic virtue: the ability to produce certain, non-subjective truth. However, over the years scientists gradually realised that mechanical representation had limitations too, as machines only record some aspects of a certain phenomenon and include their own traces in these representations. In addition, the products of these machines, such as photographs and X-rays, included a high volume of information that required interpretation to be read and understood. To solve these problems a novel paradigm emerged, which Daston and Galison name "trained judgment". This form of objectivity acknowledges scientists' expertise and experience to interpret and supplement the data provided by machines. Although Daston and Galison underline that these forms of objectivity are not mutually exclusive and can co-exist, their approach to historically study scientific practices emphasises the processes underlying the historical and situated evolution of the concept of objectivity.

A different perspective on mechanical objectivity is offered by Porter (1995), another science historian who has a more sociological approach in exploring the practices that expert communities have adopted to produce objectivity. Rather than focusing on objectivity as the ability to produce certain, non-subjective truth, Porter's analysis focuses on the ways in which objectivity is constructed by reaching consensus inside scientific communities. When this "disciplinary objectivity" (Megill 1994) is difficult to reach, weaker communities often turn to mechanical objectivity to combat distrust and secure truth. For Porter, mechanical objectivity «can never be purely mechanical» (1995: 5) and

involves instead the establishment of strict procedural rules to be followed to guarantee the objectivity of the outcomes. While most scientific communities agree that knowledge based completely on explicit rules is not fully attainable, the ideal of applying a unitary "scientific method" is often used to defend the community itself from external criticisms. For Porter, this has two main reasons: first, a decision based on quantification (or "made by the numbers") or strict protocols can remove the subjectivity of scientists, providing an appearance of impartiality and fairness; second, measurements are deemed able to help separate knowledge from its local context of production. Although Porter's work does not explore visual practices themselves, I argue that his analysis is useful to explore the shift to the quantification of visual judgment inscribed in image-processing algorithms.

In this chapter, I will not focus on the epistemic and political limitations of such a conception of scientific objectivity, which have been central to STS feminist debates (see Harding 2015). I am instead interested in exploring the work put in place by a community of experts to introduce measurable and quantifiable standards to reach objective visual assessment in a field that is perceived as highly subjective by its own members (see for instance, Fishel et al. 2020). Following Porter's conceptualisation of mechanical objectivity, in this chapter I use the case of TLIISs to explore the ongoing attempts to replace or enhance expert judgment in embryology by introducing EAS algorithms.

In popular accounts of healthcare, algorithms are often considered better than individuals at making decisions due to the amount of data they can access and reliably analyse (Watcher 2017). However, the purportedly objective nature of algorithmic judgement due to its being value-neutral has been sharply criticised by STS scholars (see Christin 2017), who have shown how the delegation of decision-making to algorithms is negotiated in practice and is often enacted as a continuation of existing practices and professionals' authority (Torenholt and Langstrup 2021). Drawing on these premises, my aim is to explore how the professional community of embryologists enacted different types of TLISs and EAS algorithms in their laboratory work, by focusing on the tensions emerging between expert and algorithmic judgment (Carlson 2018).

Some notes on methods

This chapter is based on wider research focusing on the introduction of TLIISs in IVF labs in England, which involved a variety of professional and lay stakeholders[1]. Although the perspectives of IVF patients on TLIISs have been explored in the wider project (see

[1] "Remaking the Human Body: Biomedical Imaging Technologies, Professional and Lay Visions" was a large qualitative research project (2016-2023), investigating the relationship between biomedical imaging technologies and professional and lay visions. Financial support for this research was provided by the Wellcome Trust [grant number 108577/Z/15/Z]. More information about the project can be found on the project blog available at <https://remakingthehumanbody.sbm.qmul.ac.uk/>; Last retrieved on 30/05/2022.

Hamper and Perrotta 2022) and have contributed to shape my understanding of this case, this chapter focuses on empirical materials collected to explore the perspectives of professional stakeholders only. These include over 80 hours of participant observations in two NHS fertility clinics, 43 individual interviews with a variety of IVF professionals (head of labs, embryologists, doctors, nurses, representatives of companies manufacturing TLIISs) and three focus groups with the staff of three NHS fertility clinics[2]. In addition, I have collected a large amount of documentary materials such as IVF clinic websites, promotional materials collected at fertility events such as The Fertility Show (a popular live event where perspective IVF patients can gain information about professionals and clinics) and The Fertility (the British national conference of professionals working on fertility).

As in the laboratory studies tradition (Latour and Woolgar 1979), I studied the use of TLIISs and image-processing algorithms as situated processes of knowledge production (Knorr-Cetina 1981) and as a set of epistemic practices that constitute how we know what we know (Knorr-Cetina 1999). In particular, I have studied EAS algorithms «in practice» (Christin 2017), by focusing on their use in IVF labs. To do so, I have adopted a research strategy named «algorithmic refraction» (Christin 2020) to emphasise that at the core of the analysis are the reconfigurations that occur when algorithms, people and institutions interact. In other words, my focus is not on the EAS algorithms themselves, neither in terms of their code nor the software they are embedded in, but in the way they are used in fertility labs.

Findings: automated judgment versus enhancing expert judgment

Despite the diffusion and normalisation of fertility treatment, success rates remain quite low (below 30% on average). To improve success rates, extensive research in reproductive medicine has focused on how to assess and select embryos to predict which embryos have greater pregnancy potential (see Perrotta and Giraud 2018). Conventional embryo assessment is based on morphological characteristics, that is, the appearance of the embryos and criteria such as the symmetry of cells and the presence of fragmentation[3].

[2] I am thankful to Dr Alina Geampana, one of the postdoctoral researchers working on the project, for her support in data collection with professional stakeholders, including observations and interviews. In this chapter, I am focusing only on the participant observations I have personally conducted, although similar data were collected from observations conducted by Dr Geampana in three additional sites.

[3] Fragmentation (the occurrence of small pieces of cytoplasm) is very common in human embryos, both those growing inside the body and in the IVF laboratory. In fertility treatment its assessment is crucial as higher degrees of fragmentation are associated with lower likelihood of pregnancy. The assessment of fragmentation is a typical example of potential professional disagreement, where professional judgment is required to distinguish between cells and fragments.

Due to the high level of uncertainty and the lack of certain knowledge in this field, since their introduction TLIISs have been represented as a revolutionary tool to increase the understanding of reproductive processes. Embryo videos resulting from TLIISs, in fact, do not only grant increased insight into the appearance of the embryos, but also allow the exploration of the temporal dimension of embryo development, which was previously impossible to visualise. TLIISs, in addition, provided for the first time the opportunity to generate a great amount of data without negatively affecting the embryos and their chances of success.

To analyse these data, several different EAS algorithms have been produced (see Kovacs 2014). In this chapter, I focus on the two most popular EAS algorithms and their associated TLIIS, representing two very different approaches in both their data production and their role in the embryo selection process: the «Early Embryo Viability Assessment» (henceforth Eeva) system and the «Known Implantation Data Score» (henceforth, KIDScore). Both are proprietary algorithms run through dedicated software, and their associated TLIISs generate images with very low levels of mediation[4]. However, they represent different models of accountability and judgment, which I discuss in the next two sections.

Automating judgment: The Eeva system and the pursuit of mechanical objectivity

The algorithm behind the Eeva system was originally designed and patented by a California start-up called Auxogyn (now Progyny) and then sold to the multinational pharmaceutical company Merck Serono. This algorithm was designed following the principle that there is a universal set of parameters that can predict embryo viability and, therefore, part of their marketing strategy was based on their identification of the correct set of variables which was at the base of their proprietary closed algorithm[5]. In addition, the algorithm was offered together with a TLIIS deemed able to generate data automatically (without the intervention of embryologists) and therefore usable in any lab despite different context, professional teams, and equipment. Due to these features, the Eeva algorithm was presented as able to select embryos without any human intervention based on the objective information collected. The opaque nature of Eeva and its lack of transparency were presented as a major strength of this tool, in comparison with all the other TLIIS available on the market requiring some form of human intervention to collect and analyse visual data. Eeva automated analysis generates a HIGH or LOW result per each

[4] Embryologists have to learn how to use TLIIS dishes and how to change the focus, before the system can start collecting pictures automatically. After dishes have been positioned, images are automatically collected and recorded by the software.

[5] Interestingly, before completing the divestiture of its Eeva business in January 2018 to Merck Serono, Auxogyn was at the centre of an international dispute for attempting to patent the timing of cellular development as «predictive parameters» in embryo selection (see van de Wiel 2019).

embryo (see Figure 2.1), indicating its potential without explaining how this automated judgment is produced.

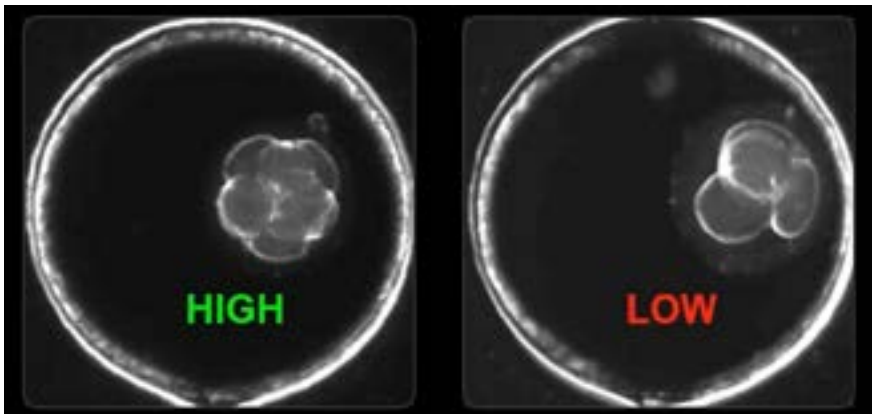

Figure 2.1 – Eeva's visual assessment[6]

Although the evidence supporting the effectiveness of the system have been contrasting (see van de Wiel 2016), the Eeva system became very popular in the UK immediately after its launch, attracting a lot of media attention when the first Eeva baby was born in Glasgow in June 2013 (see van de Wiel 2016). Just few years later, its diffusion was limited and none of the NHS clinics or the professionals interviewed were using it during the years of our fieldwork (2018-20). Several participants, however, had either used it previously or had had discussion about the system within their labs.

Interestingly, our findings show a significant resistance to the Eeva model. The selected parameters on which its automated analysis were based were part of the proprietary algorithm and therefore unknown by the professionals using it. Although in principle our participants were excited about the opportunity offered by TLIISs and algorithms (as I will discuss more in detail below), the completely opaque nature of Eeva's functioning generated significant distrust among the professional community. One of the embryologists reflected on the use of the Eeva's algorithm in their lab:

> I've never understood this system. And I feel that's why most of the people in the laboratory weren't very happy with Eeva because there is not really, or we couldn't understand the reason why the Eeva was selecting or grading the embryos in that order instead of any other orders. […] We don't trust the algorithm. The algorithm is only for like a second opinion. So if you have two embryos looking the same… But obviously if there is an embryo that is average but has got a good grade you're not going to transfer that, you're going to transfer the one that is top quality. (Embryologist)

[6] Source: "Eeva In Action (ASRM 2013)" video, available on the Eeva Test YouTube channel at <https://www.youtube.com/watch?v=hdL0kY5rPbM>; Last retrieved on 30/05/2022. The website screenshot is included in compliance with fair use guidelines, serving educational or informational purposes.

The excerpt shows how, in the lab of the interviewer, the automated judgment offered by the Eeva system ("a good grade") was rejected by the embryologists who continued making their decisions on embryo selection on the basis of their expert judgment (selecting the embryo they assessed as "top quality"). Eeva was instead used as a support, "a second opinion", when embryologists deemed that two or more embryos had the same morphological characteristics.

More broadly, critics focused on the way in which objectivity was pursued by the system. The automatic collection of data from the images removed the need to gather input by human embryologists – which is instead required by competitor systems. Eeva, instead, was able to generate outputs based on computer-based interpretations of the visual assessment of several unknown parameters. In this way, the Eeva system embedded (or aspired to embed) the type of mechanical objectivity described by Daston and Galison (2007) in their work. However, the fact that neither the parameters or the visual assessment methods were known or understood by embryologists made the system impossible to verify. In addition, the system did not just claim the ability to automatically select the relevant information to assess an embryo, but the ability to select embryos themselves. This attempt to automate judgement was sharply criticised by many of our participants. For instance, one of the senior embryologists specialised in TLIISs claimed:

> I think they are right in that… If you show the same embryo image to Eeva, it's going to give you the same parameters. Whereas if you showed it to two embryologists, they're going to have a discussion: that is definitely a fragment, no, no, I think that's a cell, etc. So the way that you bypass this subjectivity on the human side is with protocols, is with definitions and tightening them up. So they [Eeva's promoters] are correct, the machine is going to be consistent every time, the human is going to vary. But the issue is that it [Eeva] might be consistently saying something wrong. […] The consistency is there, it's the accuracy that I question. (Senior embryologist and head of embryology lab)

As the interviewee points out, the ability to guarantee consistency offered by the Eeva system was not paired with an ability to be correct, or at least to convince embryologists about its accuracy. It is important to underline that Eeva's promoters were trying to "blackbox" their criteria for predicting embryo development in a field characterised by a high level of uncertainty, where it is still widely unknown why certain embryos result in a pregnancy and others do not. As the same senior embryologist suggests, embryology is still characterised by a dominance of tacit knowledge that is embedded in the laboratory practices and reproduced inside the community, but not completely made explicit: «The embryologists do sometimes see things that a machine wouldn't know because you haven't been able to quite put it into words and if you can't define it you can't teach the machine». As I will show in the next section, being aware of what is unknown or not known yet, the community of embryologists has a variety of practices in place to limit subjectivity and reduce what they define as inter-subject variation (see Fishel et al. 2020). However, the Eeva model did not allow for these practices to be enacted.

This first closed model was rapidly dismissed when in 2018 Eeva was bought by Merck Serono, which changed the Eeva model to integrate it in their own incubator called GERI. Interestingly, the integration of Eeva into GERI was embedded in a wider transformation of the Eeva model, which opened its black box over the last few years:

> We saw the development of the way Merck presented EEVA: in the beginning they said it knows which one [embryo] to choose because they were convinced that it would choose the right ones. Then it was helping you [embryologist] to select the ones [embryos] to choose. And then in the end it was a guide. So they actually went from where they really believed that it was choosing the right things to something that would be helpful together with the normal assessment. That's why I think they were not more successful. They simply promised too much. (Company representative)

As the last excerpt shows, the wider embryology community – including both embryologists and other key actors working in the development of laboratory tools – agreed on the inability of the original closed model to make accurate automated judgement. Interestingly, even though interviewees showed a general optimism in the future improvement related to use of EAS algorithms, they also shared a wider distrust toward the idea of automated algorithmic judgement as an alternative to expert judgment made by embryologists. This distrust was summarised by the same interviewer, a representative of a competitor company: «As the technology develops, I'm pretty sure that in the future we will see even more automatisation but still we will need the final decision to be made by an embryologist».

Enhancing expert judgment: integrating the KIDScore in the co-production of objectivity

The other most popular EAS algorithm is the KIDScore, produced by Vitrolife (a company specialised in equipment and products for IVF labs) and marketed in conjunction with Embryoscope (currently the most popular TLIIS worldwide). Unlike Eeva, the KIDScore[7] required manual annotation (i.e., for an embryologist to manually register certain traits and biological events) of the most relevant characteristics of an embryo to calculate scores. These scores (from 1 to 9.9) reflect the statistical chance of implantation based on the Vitrolife database, correlating embryo development variables with implantation outcome after transfer (i.e., pregnancy or babies born) and are offered with additional information on each embryo (see Figure 2), which provide a context for their interpretation. Unlike Eeva, the KIDScore is advertised[8] as a «decision support tool»

[7] Here I am referring to the D3 or D5 versions of KIDScore, respectively for embryo selection at day 3 and day 5.

[8] Source: Evaluation tools, available on the Vitrolife website at <https://www.vitrolife.com/products/time-lapse-systems/evaluation-tools>; Last retrieved on 30/05/2022.

which is able to assign «an objective score to each embryo», but leaving professionals to decide the level of automatisation they prefer.

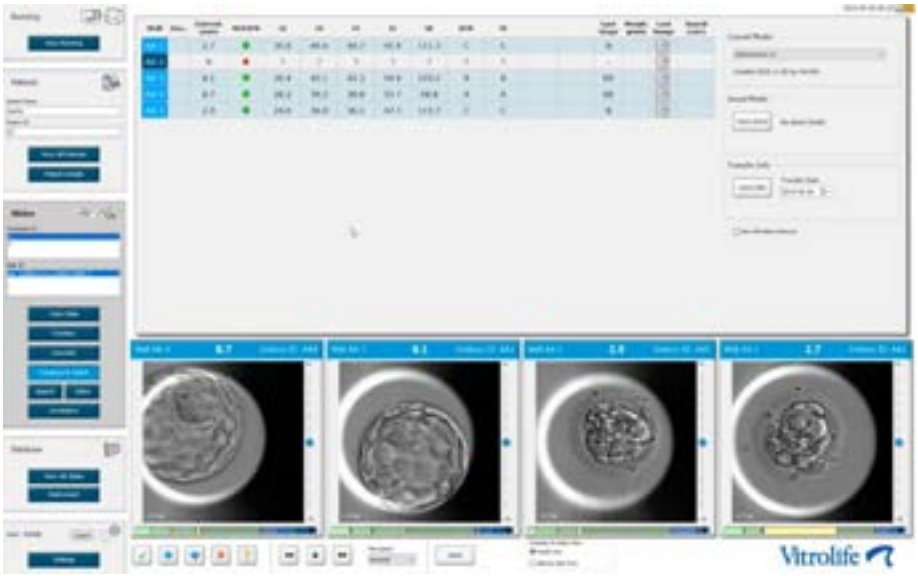

Figure 2.2 – KIDScore's visual assessment[9]

While the full automation promoted by the Eeva model encountered the resistance of large parts of the professional community, KIDScore's more open approach was welcome by many professionals. The flexibility of KIDscore consisted of allowing embryologists to select and collect data on some of the parameters suggested by the software, and even modify the algorithm by using in-house data to support the analysis instead of Vitrolife's centralised database. In this way, KIDscore allowed some space to adapt the algorithm to the needs of an individual clinic, while remaining a tool to move away from the perceived human subjectivity of the embryologists.

Subjectivity was perceived as a problem in the expert community well before the introduction of TLIISs and EAS algorithms. To generate objectivity in EAS several practices have been enacted in fertility labs to make sure that the best embryo is selected:

> I think that's [subjectivity] always going to be a problem because every embryologist is always going to be very subjective. […] We should all follow the same criteria so ideally, in an ideal world we should all be grading the same. It's never going to happen. […] Any case where we are unsure, we don't know how to grade it or it is borderline, we'll always ask a senior [embryolo-

[9] Source: "Compare and Select – optimising embryo evaluation" video, available on the Vitrolife website at <https://www.vitrolife.com/support/support-material/instruction-movies/>; Last retrieved on 30/05/2022. The website screenshot is included in compliance with fair use guidelines, serving educational or informational purposes.

gist] or a second person to come to a joint conclusion. I agree that is fair enough and objective. (Trainee embryologist)

To deal with the subjectivity of the morphological assessment, embryologists compare their judgment with others, ideally more expert, in their lab. This is a common strategy to reduce subjectivity and foster more objective outcomes and is arranged both in-house and through national quality assessment exercises (known as the UK NEQAS Reproductive Science Scheme). These exercises consist in having the same embryo assessed by some or all the members of a lab and then comparing the results to evaluate their consistency among the lab; and in the national community, activating additional training processes to guarantee uniformity if necessary.

The use of the KIDScore adds another layer of subjectivity to the process as embryologists have to record not only the morphological characteristics of the embryos but also the time of several biological events, such as the appearance of pronuclei or the cellular divisions. As mentioned earlier, this process is called "annotation" and involves an individual embryologist to manually assign a value and a time-stamp for each registered parameter in the database:

We [embryologist] can be a little subjective based on the timings or what we put as a for example, a T2, T3 [time of cellular division in 2 or 3 cells], but I think if you have a good in-house system in place that encourages uniformity amongst practitioners it will help in eliminating that factor. (Senior embryologist)

To explain what a good in-house system to foster consistency and uniformity of judgment is, the same senior embryologist provided their personal experience in joining «a good lab»:

I was shown what to do and asked to do something, i.e., annotate and grade an embryo. Once I've done that, another embryologist checks to make sure whether or not it tallies. Another embryologist will grade the same embryo in time intervals and then compare it to my result. So gradually, on a weekly basis, they see how you've progressed until there is less subjectivity between you and your trainer and then they would say "we're happy with the way you annotate, we're happy with the way you grade embryos". Then you get signed off in terms of grading on the Embryoscope. But in addition to that on a quarterly basis we will have an e-mail sent around for quality control. […] All the embryologists will write their answer on this [in-house designed] paper that has basically key information, key time intervals for T2, T4, T5, T6, T7, T8 [indicating further cell divisions] and a blastocyst [an embryo at 5]. You write next to it what time you thought these embryos have reached these stages. And then it's plotted on a spreadsheet to see the differences between us. So that gives us a control to make sure that we're all together, we're all uniform. (Senior embryologist)

This excerpt shows how the process of producing objectivity with the TLIIS follows a very similar process embryologists used in assessing morphological criteria, by putting in place lab practices that aim to limit their subjectivity and individual idiosyncrasies to

produce uniformity of results. As in Porter's (1995) analysis, objectivity is constructed by reaching consensus inside their scientific community through the establishment of strict procedural rules[10]. Although, as in Porter's case, this is not a purely mechanical form of objectivity, due to the availability of images and videos that do not negatively impact embryo development, embryologists can dedicate time to discuss their assessment criteria and practices:

> It [TLIIS] is quite good for training purposes and also to compare our grading. She [the head of the lab] could grade ones [embryos] and then I could also grade them and it wouldn't affect the embryos because they wouldn't have to come in and out. (Embryologist)

Unlike Eeva, the use of KIDScore has been integrated in established professional practices to reduce subjectivity. Although the final output of the KIDScore is to produce a score that, like in the case of Eeva, represents the embryo assessment based on a proprietary opaque algorithm, its more open approach to data production allowed embryologists to integrate its use in their local practices to enact uniformity and limit the subjectivity human judgment, which is considered part and parcel of the embryology profession. In addition, the use of the KIDSCore and, more in general, TLIISs that allow embryologists to collect their own data are seen as an opportunity to gain further information and, in the long term, develop knowledge. As one of the interviewees, a medical director involved in innovation and research in this field, stated: «In time lapse what we see is literally visualised aspects of pre-implantation development that cannot be seen in any other way».

Discussion and conclusion

Exploring the case of the introduction of TLIISs and EAS algorithms, in this chapter I analysed their use "in practice" (Christin 2017) in IVF labs, by focusing on the emerging changes in visualisation practices and different ways to use algorithms to select or support the selection of embryos.

The findings show how different relationships between image production and image processing and interpretation are enacted using different visualisation tools. As mentioned above, Eeva and KIDScore interact differently with their associated TLIIS. In both cases, the level of mediation to produce the images is minimal while the possibilities to intervene in data processing are substantially different. As I showed above, the auto-

[10] STS scholars (Cambrosio et al. 2006) define forms of objectivity based on protocols and procedures as «regulatory objectivity», to underline the endogenous nature of this form of regulation of medical communities. Although the case of the use of TLIISs represents a prime example to explore how a community tries to reach consensus, clinical practices in fertility care are significantly less standardised than in other medical fields (see Perrotta and Geampana 2021), making the case of TLIISs not representative of a wider development of regulatory objectivity in fertility care.

mated production of images themselves was not challenged. The production of objective results was shifted from image production to the processing and interpretation of images and how these were negotiated between embryologists and TLIISs. Interestingly, research participants, and particularly embryologists, seemed more interested in the potential that the TLIISs offered to visualise the early development of embryos in lab and the related knowledge production potential, rather than in the changes that occurred in visualising practices themselves. The changes required in seeing embryos through a computer monitor rather than a microscope were downplayed, while the dominant narrative focused on the possibility offered by the technology to see the embryos without disturbing them. Similarly, the transition from the assessment of still images to the assessment of videos was framed as a matter of fact without further consideration. Although a change of focus from morphology to morphokinetic in the assessment and selection of embryos is emphasised in the community and in the medical discourse (see Fishel et al. 2020), this shift is not equally discussed by professionals in terms of the changes in their visual practices. The focus is again on the possibility offered by TLIISs to observe embryos continuously and, as an obvious consequence, the need to train embryologists regarding "what" to observe, by focusing on the timing of biological events in addition to conventional appearance of the embryo. The main concerns of embryologists regarding the change of visualisation practices were related to the possibility to separate the visual embryo judgment from its context, by translating the visual elements into decontextualised data. These concerns were framed as part of more general considerations on the use of algorithms in IVF.

Furthermore, the findings show how, even when algorithms are similar – in this case, both aimed at producing (semi)automated assessment of embryo potential based on closed, unknown, proprietary algorithms – they differ significantly in the way in which are enacted in IVF practices. Despite the marketing strategies of Eeva's promoters, both algorithms were used by embryologists to support expert judgment (or provide a second algorithmic opinions) rather than as an automated alternative to it. Interestingly, several interviewees claimed that the inability of algorithms to replace expert judgment was probably due to the fact that most of the current limited knowledge in embryology is still tacit and not yet made explicit. In addition, there was a consensus on the fact that the expert judgment of a human embryologist will be needed also in the foreseeable future. The main reason for this lack of trust toward algorithms seems to lie in what Ziewitz (2016) calls their "inscrutability". Current EAS algorithms, with their closed, unknown, proprietary structure, show the same lack of transparency and the limited accountability of algorithms discussed in the literature (Pasquale 2015). These characteristics, though, do not allow embryologists to integrate them in their lab practices and require a leap of faith in the automated judgement that professionals are not willing to take. However, the new possibilities of visualisation offered by the use of TLIISs allow embryologists to improve their «disciplinary objectivity» (Megill 1994) by integrating them in their practices to limit subjectivity. Similarly to what is already suggested in the literature (Torenholt

and Langstrup 2021), the use of EAS algorithms in practice does not reflect the expected disruption proposed by Eeva's promoters but rather a continuation of existing practices aimed at co-producing objectivity through the interactions of embryologists and more sophisticated machines. For these reasons, the more flexible structure of the KIDScore makes this algorithm more appealing to the expert community – although further attempts to automate judgment are made by TLIISs' manufacturers[11].

Acknowledgements

I presented an early version of this chapter as an invited contribution to a workshop organised by Marina Maestrutti in Paris in March 2017 titled "Voir savoir et donner à voir". I would like to thank Marina and all the other participants for the useful feedback. I am also grateful to the editors of this book, Simone Arnaldi, Stefano Crabu and Assunta Viteritti, and all my collaborators in the "Remaking the Human Body Project". I thank Elizabeth Mason for proofreading the chapter and all the research participants for their time and interest in the project. Financial support for this research was provided by the Wellcome Trust, United Kingdom [108577/Z/15/Z].

[11] Vitrolife has recently introduced a new tool called IDAScore (an AI-based algorithm with fully automated scoring of embryos) which follows Eeva's principle. However, as this tool did not exist at the time of empirical work and it is not very diffused yet, I have not discussed this in this chapter.

Alac, M.
2008 'Working with Brain Scans: Digital Images and Gestural Interaction in fMRI Laboratory', *Social Studies of Science*, *38*(4), 483-508.

Ananny, M.
2016 'Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness', *Science, Technology, & Human Values*, *41*(1), 93-117.

Beaulieu, A.
2002 'Images Are Not the (Only) Truth: Brain Mapping, Visual Knowledge, and Iconoclasm', *Science, Technology, & Human Values*, *27*(1), 53-86.

Beaulieu, A., De Rijcke, S., Van Heur, B.
2012 *Authority and Expertise in New Sites of Knowledge Production*, in P. Wouters, A. Beaulieu, A. Scharnhorst, S. Wyatt (eds.), *Virtual Knowledge: Experimenting in the Humanities and the Social Sciences* (pp. 25-56), Cambridge (MA), The MIT Press.

Burrell, J.
2016 'How the machine "thinks": Understanding opacity in machine learning algorithms', *Big Data & Society*, *3*(1), 1-12.

Burri, R. V.
2008 'Doing Distinctions Boundary Work and Symbolic Capital in Radiology', *Social Studies of Science*, *38*(1), 35-62.

Cambrosio, A., Keating, P., Schlich, T., Weisz, G.
2006 'Regulatory objectivity and the generation and management of evidence in medicine', *Social Science & Medicine*, *63*(1), 189-199.

Carlson, M.
2018 'Automating judgment? Algorithmic judgment, news knowledge, and journalistic professionalism', *New Media & Society*, *20*(5), 1755-1772.

Christin, A.
2016 'From daguerreotypes to algorithms: Machines, expertise, and three forms of objectivity', *ACM Sigcas Computers and Society*, *46*(1), 27-32.

Christin, A.
2017 'Algorithms in practice: Comparing web journalism and criminal justice', *Big Data & Society*, *4*(2), 2053951717718855.

Christin, A.
2020 'The ethnographer and the algorithm: Beyond the black box', *Theory and Society*, *49*(5), 897-918.

Coopmans, C., Vertesi, J., Lynch, M. E., Woolgar, S. (eds.)
2014 *Representation in scientific practice revisited,* Cambridge (MA), The MIT Press.

Daston, L., Galison, P.
2007 *Objectivity,* Princeton (NJ), Princeton University Press.

De Rijcke, S., Beaulieu, A.
2014 *Networked Neuroscience: Brain Scans and Visual Knowing at the Intersection of Atlases and Databases*, in C. Coopmans, J. Vertesi, M. E. Lynch, S. Woolgar (eds.), *Representation in Scientific Practice Revisited* (pp. 131-152), Cambridge (MA), The MIT Press.

Fishel, S., Campbell, A., Foad, F., Davies, L., Best, L., Davis, N., Smith, R., Duffy, S., Wheat, S., Montgomery, S., Wachter, A.
2020 'Evolution of embryo selection for IVF from subjective morphology assessment to objective time-lapse algorithms improves chance of live birth', *Reproductive Biomedicine Online*, *40*(1), 61-70.

Geampana, A., Perrotta, M.
2021 'Predicting Success in the Embryology Lab: The Use of Algorithmic Technologies in Knowledge Production', *Science, Technology, & Human Values*, *48*(1), 212-233.

Gillespie, T.
2014 *The Relevance of Algorithms*, in T. Gillespie, P. J. Boczkowski, K. A. Foot (eds.), *Media Technologies: Essays on Communication, Materiality, and Society* (pp. 167-194), Cambridge (MA), The MIT Press.

Hamper, J., Perrotta, M.
2022 Watching embryos: Exploring the geographies of assisted reproduction through encounters with embryo imaging technologies, *Social & Cultural Geography*, *24*(9), 1557-1575.

Harding, S.
2015 *Objectivity and Diversity,* Chicago (IL), University of Chicago Press.

Joyce, K.
2005 'Appealing Images: Magnetic Resonance Imaging and the Production of Authoritative Knowledge', *Social Studies of Science*, *35*(3), 437-462.

Joyce, K. A.
2006 'From Numbers to Pictures: The Development of Magnetic Resonance Imaging and the Visual Turn in Medicine', *Science as Culture*, *15*(1), 1-22.

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

Knorr-Cetina, K.
1999 *Epistemic Cultures: How the Sciences Make Knowledge*, Cambridge (MA), Harvard University Press.

Kovacs, P.
2014 'Embryo selection: The role of time-lapse monitoring', *Reproductive Biology and Endocrinology*, *12*:124.

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

Lynch, M., Woolgar, S. (eds.)
1990 *Representation in Scientific Practice*, Cambridge (MA), The MIT Press.

Maestrutti, M.
2008 'Rendere visibile l'invisibile: Il ruolo della visione e dell'immagine nella costruzione di una storia delle nanotecnologie', *Rassegna Italiana di Sociologia*, *49*(3), 427-450.

Maestrutti, M.
2011 *Imaginaires des nanotechnologies: Mythes et fictions de l'infiniment petit*, Paris, Vuibert.

Megill, A. (ed.)
1994 *Rethinking objectivity*, Durnham, Duke University Press.

Pasquale, F. A.
2015 *The Black Box Society: The Secret Algorithms That Control Money and Information,* Cambridge (MA), Harvard University Press.

Perrotta, M.
2013 'The Study of Technoscientific Imaging in STS', *TECNOSCIENZA: Italian Journal of Science & Technology Studies*, *3*(2), 163-176.

Perrotta, M., Geampana, A.
2020 'The trouble with IVF and randomized control trials: Professional legitimation narratives on time-lapse imaging and evidence-informed care', *Social Science & Medicine*, *258*, 113115.

Perrotta, M., Geampana, A.
2021 'Enacting evidence-based medicine in fertility care: Tensions between commercialization and knowledge standardization', *Sociology of Health & Illness*, *43*(9), 2015-2030.

Perrotta, M., Giraud A. S.
2018 'Enacting standards in imperfect worlds: Exploring flexibility in practicing standards', Paper presented at *SCOS/ACSCOS Conference – Imperfection, incompleteness and impermanence in organizational life*, 17-20th August, Tokyo, Japan.

Porter, T. M.
1995 *Trust in numbers: The pursuit of objectivity in science and public life*, Princeton (NJ), Princeton University Press.

Prasad, A.
2005 'Making images/making bodies: Visibility and disciplining through magnetic resonance imaging (MRI)', *Science, Technology & Human Values*, *30*(2), 291-316.

Prasad, A.
2007 'The (Amorphous) Anatomy of an Invention: The Case of Magnetic Resonance Imaging (MRI)', *Social Studies of Science*, *37*(4), 533-560.

Reich, A.
2012 'Disciplined doctors: The electronic medical record and physicians' changing relationship to medical knowledge', *Social Science & Medicine*, *74*(7), 1021-1028.

Roepstorff, A.
2007 *Navigating the Brainscape: When Knowing Becomes Seeing*, in C. Grasseni (ed.), *Skilled Visions: Between Apprenticeship and Standards* (pp. 191-206), New York & Oxford, Berghahn Books.

Torenholt, R., Langstrup, H.
2021 'Between a logic of disruption and a logic of continuation: Negotiating the legitimacy of algorithms used in automated clinical decision-making', *Health*, *27*(1), 41-59.

Van de Wiel, L.
2016 *Cellular Origins: A Visual Analysis of Time-Lapse Embryo Imaging*, in M. Lie & N. Lykke (eds.), *Assisted Reproduction Across Borders* (pp. 306-319), New York, Routledge.

Van de Wiel, L.
2019 'The datafication of reproduction: Time-lapse embryo imaging and the commercialisation of IVF', *Sociology of Health and Illness*, *41*(S1), 193-209.

Wachter, R.
2017 *The Digital Doctor: Hope, Hype, and Harm at the Dawn of Medicine's Computer Age*, New York, McGraw-Hill Education.

Zarsky, T.
2016 'The Trouble with Algorithmic Decisions: An Analytic Roadmap to Examine Efficiency and Fairness in Automated and Opaque Decision Making', *Science, Technology, & Human Values*, *41*(1), 118-132.

Ziewitz, M.
2016 'Governing Algorithms: Myth, Mess, and Methods', *Science, Technology, & Human Values*, *41*(1), 3-16.