OPEN ACCESS

University of BRISTOL

Nunes Vieira, L., Ragni, V., & Alonso, E. (2021). Translator autonomy in the age of behavioural data. *Translation, Cognition & Behavior*. https://doi.org/10.1075/tcb.00052.nun

Peer reviewed version

Link to published version (if available):
10.1075/tcb.00052.nun

Link to publication record in Explore Bristol Research
PDF-document

University of Bristol - Explore Bristol Research
General rights

Translator autonomy in the age of behavioural data

Lucas Nunes Vieira[1], Valentina Ragni[1] & Elisa Alonso[2]
[1]University of Bristol, United Kingdom | [2]Universidad Pablo de Olavide, Spain

Translation behaviour is increasingly tracked to benchmark productivity, to calculate pay or to automate project management decisions. Although in many cases these practices are commonplace, their effects are surprisingly under-researched. This article investigates the consequences of activity tracking in commercial translation. It reports on a series of focus-group interviews involving sixteen translators who used productivity tools to independently monitor their work for a period of sixteen weeks. Our analysis revealed several ways in which the act of tracking activity can itself influence translators' working practices. We examine translators' conceptualisations of productivity and discuss the findings as a matter of translator autonomy. The article calls for further awareness of individual and collective consequences of monitoring translation behaviour. Although in some contexts translators found activity tracking to be useful, we argue that client-controlled tracking and translator autonomy are in most cases incompatible.

Keywords: translator autonomy; translation productivity; behavioural data; activity tracking; translation technology; CAT tools.

1. Introduction

Technology plays an increasingly important role in professional translation. Computer-assisted translation (CAT) tools streamline translators' work by allowing them to translate a range of file types while leveraging content from translation memories. When translation memory matches are not available, machine translation (MT) can be used as a source of assistance. CAT tools are also important for project and terminology management. They simplify collaborative work by ensuring consistency and communication within and between teams, which is often facilitated by cloud or dedicated server technologies.

Among their many benefits, translation technologies are usually expected to have some type of positive impact on translation productivity. The online community for users of CAT tool Trados Studio is, for instance, called just that: 'translation productivity' (RWS n.d.). Similarly, a blog post by developers of memoQ, another well-known CAT tool, refers to a 'productivity boost' in describing the tool's benefits (Starkmann 2020). Productivity is a complex concept, but its meaning on these websites is implied: the functionalities of a CAT tool are expected to allow translators to deliver larger textual volumes per unit of time while keeping product quality unaffected or ideally while improving it. Speed and other aspects of the translation process, such as amount of typing and editing, are consequently key variables in technologized workflows. Together with product assessments, translators' behaviour is what confirms translation technologies' usefulness. Speed and other types of work quantification may also inform project management by showing on the fly how a project is progressing. Furthermore, the time translators spend on a project may be used to calculate pay and the return on investment of any resources.

It should not come as a surprise, therefore, that language service providers started to collect translators' behavioural data, including their speed and the number of edits they perform. The *Dynamic Quality Framework* (DQF) of the Translation Automation User Society (TAUS), for instance, can be integrated into CAT tools to pool data on translation speed and other variables to enable "real-time reporting" (TAUS 2020). Translation companies can then use this data to benchmark their performance against industry averages, which might allow them to diagnose problems with their MT systems though also to evaluate their translators relative to the remainder of the industry. Cloud-

based platforms facilitate this type of tracking and analysis. Previous research has called attention to how companies like Straker, for instance, make use of behavioural data in the cloud (García 2017). At the time of writing, Straker addresses prospective clients on their website by describing a job assignment system that uses data analytics "to know which translators will go faster on your content but produce excellent quality"; according to their website, they can also establish "the normal rate a translator will work on this type of content and language pair to predict time-frames" (Straker n.d.). In relation to behavioural payment models, MT-based translation provider Unbabel, for example, explains to their bilingual, though not necessarily professional, translators: "Your hourly rate is calculated on a daily basis by an algorithm that analyses your speed and quality"; they add: "as a very rough guide you'll need to be completing average-length emails of 200-300 words within five minutes to access the upper hourly rates" (Bartlett 2020).

Although not all companies that use tracking information do so in the same way or to the same extent, the examples above and the data collection capabilities of initiatives like the DQF show how using behavioural data to inform or automate decisions is now a reality of the language industry. In the more extreme cases, algorithms are used to triage commissions and to profile translators based on how fast they work, which can have consequences for pay and the likelihood of receiving work in the future. Importantly, these practices are being implemented without a robust understanding of their long-term implications. While in recent years there has been growing discussion of ethical issues, business models and professional standards in commercial translation (e.g., García 2017; LeBlanc 2017; Moorkens 2020), translators' perspectives on current uses of behavioural data have received less attention. In addition, the acts of tracking activity and measuring productivity might themselves have implications for how translators work. Observing an activity often changes it—a widely studied phenomenon that can be referred to as the *observer effect*[1] (e.g. Baclawski 2018, 83)—so the use of logging tools to observe the translation process may generate or alter behaviours, which has not, to our knowledge, been systematically researched in professional translation either.

In what follows, we present a qualitative study that seeks to highlight the consequences of monitoring translation processes in professional contexts. As part of the study, sixteen translators, seven in-house and nine freelance, used productivity tools to observe their own work. We flipped the paradigm where translators are tracked in the cloud to one where they independently keep track of their own working practices. Translators provided us with summaries of their activity and with perceptions of the process of tracking it. They reported to us once a week for sixteen weeks. We then followed up on the weekly reporting with a series of focus-group interviews. Collecting translators' views across time and interviewing them at the end of the process gave us valuable insight into any changes in perception or in how translators dealt with specific issues. The study draws on this analysis to fulfil two goals. First, based on translators' own assessment, the study aims to explore how the act of tracking activity might itself affect translators' work. Second, it aims to discuss the implications of using behavioural data in professional translation, including the potential ethical concerns raised by this practice and how it may have consequences for translators' professional standing, which we discuss as a matter of autonomy.

In the remainder of the article, we provide a brief literature review in § 2, where we discuss the concept of autonomy and previous research in translation and other areas linked to surveillance and work quantification. We provide details of our methodology in § 3 and present our results in § 4. The results we report here are based primarily on the focus-group interviews rather than on translators' interim self-reporting, which we cover in more detail in a separate industry-oriented report (Ragni, Alonso & Vieira, forthcoming). Finally, we discuss implications of the results in § 5 and conclude the article in § 6.

2. Literature review

2.1. *Translator autonomy*
Autonomy is an important concept in discussions of privacy, surveillance, and artificial intelligence. Individual autonomy is often associated with Kantian moral philosophy, but the concept in fact

---

[1] Other formulations include the *observer's paradox* (Labov 1972, 209) or, less accurately as a term to use in the social sciences, the *Heisenberg effect* (see Salkind 2010).

predates Kant by a considerable margin (Swaine 2016). *Autonomy* means acting in a way that is not 'the product of manipulative or distorting external forces' (Christman 2020, n.p.); it means 'living by one's own laws' (Swaine 2016, 217). Autonomy is closely connected to other concepts used to discuss translators' professional influence on their environment, most notably the concept of agency. In translation studies, *agency* is defined as translators' 'willingness and ability to act', where *willingness* relates to translators' 'consciousness, reflexivity and intentionality' (Kinnunen & Koskinen 2010, 6). The 'willingness' component of the *agency* definition can therefore be considered to—at least, in part—concern matters of autonomy. Indeed, some degree of overlap between *agency* and *autonomy* is widely recognised, for example in frameworks that seek to ascribe these properties to non-human entities (Luck & d'Inverno 1995). In some of these frameworks, *agency* presupposes *autonomy* (Franklin & Graesser 1997, 24). In others, *autonomy* transcends *agency* by referring not just to actions but also to the agent's ability to pursue their own agenda according to their own motivations (Luck & d'Inverno 1995, 258).

In an analysis of the work of Foucault including, of particular relevance to our discussion in this article, *Surveiller et punir* (*Discipline and Punish,* Foucault 1975), *autonomy* transcends *agency* so radically that the argument in fact rejects the possibility of individual autonomy because this would also require rejecting the influence of a broader form of societal power that any individual is necessarily subject to (Bevir 1999, 66). While recognising these wider societal influences, our approach to autonomy in this article is aligned with those conceptualisations of the term that emphasise self-motivated behaviour (Luck & d'Inverno 1995, 258). Our understanding of *autonomy* is therefore akin to formulations of *agency* based on notions of 'willingness' (Kinnunen & Koskinen 2010, 6; see also Buzelin 2011, 6). We focus on autonomy rather than agency to foreground a level of self-determination that would not normally be associated with translation memory tools, for instance. While translation tools can be considered agents (e.g., Olohan 2011), at present they are arguably incapable of living by their own laws (Swaine 2016, 217) or of pursuing their own self-motivated agenda (Luck & d'Inverno 1995, 258). We therefore explore the implications of activity tracking for translators' status as *autonomous* agents: their ability and intention to define the logic that governs their own behaviour according to their individual and collective professional interests.

2.2. *Activity tracking in translation and other uses of behavioural data*

In academic contexts, it is common practice to use behavioural data to gain insight into the translation process. Research in this area has used methods such as eye tracking (O'Brien 2011), electroencephalography (Hansen-Schirra 2017), brain imaging techniques (Zheng et al 2020) and, more commonly, time measurements (Toral, Wieling & Way 2018). While research can influence industry practices—and vice versa—behaviour is not normally measured commercially or for profit in these studies, but rather to improve knowledge and understanding. We therefore do not consider previous research of this nature to be part of what we analyse in this paper.

Regarding translators' attitudes to activity tracking in commercial contexts, there is some evidence that this divides opinion. Some translators may find it beneficial in specific cases whereas others feel that, depending on the tracking method, behavioural data does not capture the actual amount of work invested in translation tasks (Vieira & Alonso 2018, 16). It has also been argued that, if security and privacy concerns are addressed, activity tracking can be beneficial for translators (Moran 2018). Moran mentions how sharing productivity data might allow translation companies to intervene in the translation process by providing improved MT systems and making other technological investments (2018, n.p.).

More broadly, behavioural data collection is a societal phenomenon that transcends the translation industry. The accounting firm PwC, for instance, has come under criticism recently for developing a webcam face recognition system to check if workers were at their desk when they had to work from home during the Covid-19 pandemic (Webber 2020). On the other hand, tracking may also be prompted by a need to protect workers' free time. In 2019, a European court ruling required companies in the European Union to track employees' working hours as a way of protecting breaks and downtime (Bershidsky 2019). Beyond the world of work, businesses are increasingly aware of the information hidden in behavioural patterns harnessed by machine learning in models that predict what individuals are likely to do, how they are likely to spend their money or who they are likely to vote for. Harvesting this type of information for profit has been referred to as *surveillance capitalism* (Zuboff 2019). Concepts such as the 'internet of things' (Ashton 2009)—which envisage a reality

where daily activities such as driving or physical exercise may be recorded online at all times—also illustrate the pervasive potential of behavioural or activity tracking. Surveillance in this context concerns everyday life and not necessarily professional behaviour. There are nevertheless parallels between surveillance capitalism and activity tracking in translation, especially where the tracking of behaviour can financially benefit the client or tracking party (see also O'Neil 2016). For our purposes in this article, we deem surveillance to involve behavioural tracking that is not under translators' direct control (see § 4.2).

In work settings, part of the rationale for tracking activity is linked to what some industries call *gamification*. This is a process whereby the act of tracking behaviour is seen as a "fun" challenge that prompts individuals to improve their metrics based on some type of scoring system (Zuboff 2019, 216). In translation, this would mean using activity tracking as an incentive to maintain or improve productivity. Gamification is not dissimilar from the logic underlying uses of behavioural data like those described in § 1. If efficiency is rewarded with more work or higher levels of compensation, this may in principle motivate translators to become faster and gamify aspects of their behaviour such as speed and keyboarding. However, in the case of complex activities where the task cannot be easily quantified, organisational behaviour research in fact describes this logic as counterproductive (Ranganathan & Benson 2020, 599). Moreover, the diverse nature of translation services may further complicate assumptions around the benefits of activity tracking. This is the case at least for two reasons. First, the central role of translation technologies in the language industry may prompt misconceived perceptions of translation as a repetitive and mentally undemanding task (do Carmo 2020). This in turn may help to characterise translation as easily quantifiable in seconds or words, which may apply to some tasks or texts but not others. Second, and perhaps more gravely, translation is a predominantly freelance profession (e.g., EC, CIOL, and ITI 2017). This means that tracking and quantifying work may have different implications for translation compared to other industries that rely predominantly on salaried employment.

Although, as mentioned, translation research is attuned to the increasing use of tracking methodologies (e.g., García 2017), previous research has rarely provided an empirical analysis of translators' perceptions of activity tracking, other than measuring productivity based on strict assumptions or discussing professional standards and the role of technologies. Below we outline the profile of translators who took part in our study and how they used CAT-tool features to keep track of their working patterns.

3. Methodology

3.1. *Translators*

The seven in-house translators who took part in the study were employed by a translation company in Spain with 10-15 employees. Most of the nine freelance translators in the study lived in the UK, but this group also included translators based in Italy, Spain, Germany and China. Across the full sample, professional experience as a translator ranged between 1 and 22 years (mean = 9). Language pairs also varied. Those in the in-house group translated from a variety of European languages (e.g., English, German, French) into Spanish. Those in the freelance group worked with English, French, German, Italian, Spanish, Russian and Chinese as either a source or target language. Similarly, areas of specialisation spanned several domains, notably technical texts, games, fashion, marketing, law and business.

The in-house translators were recruited for the research through their company. Those in the freelance group were recruited largely through the local network of a professional association for translators in the UK, though some were recruited through the authors' professional contacts. As part of the process of inviting interested translators to participate in the project, we held a preliminary workshop for members of the network in the UK to introduce the study and current productivity measurement practices undertaken in the industry (e.g., in relation to existing tools and benchmarking practices like those of TAUS). All translators received full details of the study before agreeing to take part. The productivity information provided prior to the focus-group interviews (see § 3.2) was in all cases self-reported and did not involve sharing any texts or translations, but rather only information that could help to contextualise translators' perceptions, such as word counts, domains (e.g., technical or legal) and self-recorded speed. The study was approved by the research

ethics committee at the first author's institution and translators renewed their consent at different points during the investigation.

In addition to being a professional translator, being familiar with CAT tools and using them regularly in their work were also criteria for taking part. This was due to our research interest in activity tracking, which is often embedded in CAT tools or available through CAT-tool plugins. Part of the weekly self-reported data was based on activity reports translators obtained through productivity features of memoQ and Trados Studio (see § 3.2). All translators were therefore users of at least one of these tools.

3.2. *Data collection*

Data was collected in two phases: a cross-temporal phase of weekly self-reporting and a final phase, where each translator took part in a one-off focus-group interview. The cross-temporal data collection took place in the first half of 2020 and lasted sixteen weeks. The exact start dates were staggered depending on when each translator joined the study. A total of sixteen rolling weeks of self-reporting applied in all cases, so those who started later also finished later. The focus-group interviews took place shortly afterwards, when translators were able to reflect on the experience of tracking their activity in the previous weeks. Although for context we describe the study's two phases of data collection below, as mentioned in § 1, this article concentrates on the focus-group interviews (see § 4). We also note that as a "live" study, our focus was not on methodological control, but rather on translators' real-world experience of tracking their activity. This means we do not seek to provide generalisable quantitative findings, but instead to discuss implications of activity tracking based on translators' empirical perceptions.

3.2.1 *Self-monitoring of activity*. The first step in taking part in the study was to set up tools that the translators could use to monitor their work. We provided supporting information on two tools that they could use for this purpose, namely memoQ's in-built timer (memoQ n.d.) and the Qualitivity plugin for Trados Studio (RWS Community n.d.). Translators relied on these tools in addition to any methods that they were already using to track productivity, including online screen timers like Clockify.me. They then reported to us each week with information about their volume of work and speed, and their perceptions. They provided the information using an online form and a spreadsheet.

For each week of the project, the spreadsheet asked translators to report word counts, number of words translated per unit of time, subject area, language pair, type of task (e.g., if translation or proofreading) and if they had used machine translation and other technologies such as term bases or translation memories. The information was provided per project, per week. Freelance translators were also asked to mention any tasks that could not be broadly classed as *translating* and that would not therefore be captured by their CAT tool. These could include completing forms, preliminary research, terminology management or any administrative task which, in their own assessment, was not part of what they were directly paid for. We did not include this question in the spreadsheets completed by in-house translators because, as employed rather than self-employed professionals, their roles within the company by default included a wide range of tasks, such as managing projects or interns.

On the online form, translators rated how satisfied they were with their productivity that week and how useful they thought the activity data was. The online form asked translators to mention how they set up the productivity measurement features each week and any new settings they might have decided to use. The form also included open-text questions that asked translators to elaborate on their ratings and provide any comments they wished to make that week about the experience of recording their own activity.

3.2.2. *Focus-group interviews.* Once the weekly self-reporting was complete, we carried out five focus-group interviews with translators in June of 2020. Although a single researcher took the lead in each interview, all interviews included two members of the research team who worked in tandem. Before the focus groups, the in-house translators' company director was also interviewed to provide context about the company. We interviewed the in-house translators in two separate groups with four and three translators each. The freelance translators were interviewed in three groups of three. The interview durations ranged between 1h24min and 1h50min.

The interviews were semi-structured (Adams 2015). This means that while we had a range of topics to cover—e.g., the impact of Covid-19 and translators' conceptualisations of productivity (see § 4)—we also allowed unforeseen topics to emerge. The semi-structured nature of the interviews also meant that translators could comment on experiences with activity tracking they might have had outside of the study, which complemented our analysis, especially in relation to the study's second goal (see § 1). Before the focus groups, we carried out a preliminary analysis of the weekly submissions, which informed the interviews and allowed us to ask translators to elaborate on any issues they had mentioned. The interviews were conducted and recorded online in Microsoft Teams, and then transcribed by a research assistant. We used the RQDA R package (Huang 2016) to qualitatively code the interview transcripts. The codes largely followed the interview topics that emerged from translators' weekly submissions. We used the codes as a way of documenting the analysis and not as a method of quantifying it, so the coding categories could overlap and be applied to any stretch of transcript without strict coding units (e.g., words, paragraphs or sentences). The first author preliminarily coded all transcripts and a second member of the team cross-checked the full material. The coding categories are presented in § 4.

4. Results

The information translators provided in their weekly submissions reflected a wide range of contexts and experiences. Since the study took place in the first half of 2020, translators' weekly self-reporting was under way when the Covid-19 pandemic first hit Europe with national lockdowns and government instructions for the population to work from home. Given the potential impact of those circumstances on translators, this became mid-study an important factor to consider, which we did mainly through the focus-group interviews. Through the interviews, translators were able to reflect on the first phase of the study with hindsight, which allowed them to comment on whether and to what extent using the productivity tools influenced their work. Table 1 presents the coding categories we used to document our analysis.

Table 1. Coding categories used in the analysis of focus-group interviews, in alphabetical order.

Categories used for qualitative coding	
Affect	feelings about productivity and the process of measuring it
Conceptualising productivity	understandings of productivity and its influencing factors
Covid-19	any impacts of the Covid-19 pandemic on translators' work
Effects of tracking	whether tracking itself affected translators' work and how
Expertise-time-compensation	intricate relationships between expertise, time, and compensation
Invisible tasks	tasks that behavioural data can misrepresent
Internal vs external tracking	distinguishing between tracking that is controlled by translators and tracking that is controlled by an external party
Profitability	any comments on profitability, earnings and earning power
Project management	any aspects of managing a project or of the role of behavioural data in project management
Surveillance	ways in which use of behavioural data can be conceptualised as surveillance
Technical	technical issues encountered with the productivity tools

While the categories were interconnected and in one way or another all of them underlie the issues discussed, we group the findings below in four sections that are more directly concerned with *Effects of tracking* (§ 4.1); *Internal vs external tracking* and *Surveillance* (§ 4.2); *Invisible tasks* (§ 4.3); and *Profitability* and *Conceptualising productivity* (§ 4.4).

4.1. *How activity tracking affects behaviour*

Although we do not provide a quantitative analysis of the impacts of activity tracking on translation speed, according to translators' assessment, the process of monitoring their behaviour mostly corroborated rather than improved perceptions of their own speed or throughput: "I will say that [the assessment of my productivity] has stayed pretty much the same" (IH05); "it confirmed my perceived productivity I had before the study" (FL11); "I was never surprised [by] my productivity […] this is what I imagined" (FL05).[2] The corroborating role of the data was nonetheless found to be helpful. Translators thought it was useful to confirm how their working pace tended to vary according to factors such as language pair, familiarity with the topic and the extent to which they enjoy reading and working with the content. In some cases, translators were also surprised by their own working patterns such as number of breaks, how profitable their work is (see § 4.4) or simply the amount of work they undertake. As mentioned, these surprising factors did not usually or directly concern speed.

There were, however, a few cases where translators reported feeling motivated to avoid interruptions and keep a steady—and possibly faster—pace because they were logging their activity. Those who reported this described it as an attempt to beat their own times, much like the *gamification* phenomenon mentioned in the organisational behaviour literature (see § 2):[3] "you can kind of play a game with yourself and try to beat yourself" (IH03); "I was trying to beat my words per minute rate" (FL07). There are three aspects of gamification that are important to highlight based on our data. First, gamification can have negative consequences. Translator FL07 reported producing work of lower quality when trying to use the productivity measurement as an incentive to beat their own time. In this case, therefore, the gamification effect was short-lived and not sustainable. Second, gamification is not necessarily induced by tracking tools. Another freelance translator reported regularly adopting gamification techniques by setting interim hourly targets, for instance. This was described positively, but it was an old habit that could in principle be done organically without any logging software. Third, on occasion, gamification simply did not have the expected result and could in fact be demotivating when productivity was not perceived to improve: "I also tried to beat my word per minute rate, but I wasn't very good at it and it was kind of disappointing when the next week I saw that it was even lower than the last week" (FL12). Translators also mentioned that interpreting the information generated by the tools was on occasion difficult because of the complexity and granularity of the data. Issues with interpretability may therefore have helped to undermine any potential positive effects of gamification in their assessment.

Notably, it was mentioned that activity tracking can prompt behaviours that aim to work around misconceptions or problems in the activity tracking method. A translator mentioned, for instance, a previous proofreading job where the translation company would keep track of the number of corrections and challenge translators for passages left unamended. When certain sections did not require any amendments, the translator reported at times simply deleting and reinstating a full stop just to flag that the lack of amendments was intentional. Behaviours of this nature are not inherently caused by the translation task, but rather by a need to address commissioners' potential misapprehensions of what counts as desirable practices. If client misconceptions are combined with the ability to track what translators do, this can therefore prompt what we call *artefact behaviours*. These behaviours react to a logic where productivity is sought based on flawed proxy measures: "you sort of almost falsify the result to show that you were active […] it's productivity for productivity's sake, as opposed to actually being useful or creating quality" (FL01). Similarly, this type of behaviour can also be prompted by the observer effect itself: "if you know that somebody is looking over your shoulder I think it damages creativity […] you're suddenly sort of double-checking, you know, whether you're too slow, too fast, whether you're doing it right" (FL02). It is important, therefore, to distinguish between gamification and artefact behaviours. Gamification—even if ineffective—is prompted by a

---

[2] Participant 5, in-house group (IH). Participant ID codes follow this same pattern in the remainder of the article, where FL stands for those in the freelance group.

[3] *Auto-gamification* may in fact be used in cases where the gamification process is "inadvertent" and occurs at an "individual-level" (Ranganathan & Benson 2020, 573). Here we use simply *gamification* because our data suggested this was sometimes a conscious decision on the part of translators and therefore not necessarily something inadvertent. *Gamification* was also preferred to avoid a potential confusion with automated aspects of the translation process; for example, concerning the use of algorithms or MT.

desire to improve actual productivity and does not necessarily involve an external observer, since it often implies competing with oneself. Artefact behaviours, by contrast, are artificial and stem from the prospect of an external observer and their expectations or potential misunderstandings of the translation process. Artefact behaviours risk running counter to some of the assumed benefits of tracking the translation process in the first place.

4.2. *Internal vs external tracking*

As mentioned in § 1, behavioural data can be used to calculate compensation. Translators mentioned different payment models they knew of that involved being paid proportionally to the amount of typing they carried out or the amount of time they spent translating. In situations where they were paid hourly rather than per word, it was largely uncontroversial for translators in the study that keeping track of how much time one spends on a task is important. However, they also mentioned situations where the time measurement was embedded in editing interfaces controlled by the client, which raised concerns about practices perceived to be invasive.

Who decides to monitor activity is therefore a crucial factor in how beneficial this monitoring is for translators. Some of the potential benefits of activity tracking mentioned in the study (see § 4.4) largely apply to circumstances where translators are in control of the measurement. There are therefore important differences between *internal* tracking, when translators control the monitoring of their own activity within their company or business, and *external* tracking, when monitoring is controlled by the client or other stakeholders involved in the commission: "For me, there's quite a lot of pros owning it [productivity tools] as an individual […] all the cons come when you're sharing the data with someone else as a mandatory thing" (FL01). Specifically, the drawbacks of external tracking were often associated with the possibility that clients can take financial advantage of the information: "I think clients are there obviously to make a higher margin and I am quite sure that if they could have access to our productivity data, they would probably use it […] to drive our rates down" (FL11). This may happen, for instance, if translators are faster than expected, which may be down to their expertise and experience rather than an "easier" task. Translators also raised privacy concerns and mentioned how sharing productivity data with clients can undermine their personal and professional freedom: "I don't like people having the opportunity to judge me for going for a walk […] as long as I deliver the job, at a decent price, and the work is good, who cares?" (FL07). Translators in our focus groups associated external monitoring practices with employees having to clock in and out of work. The external tracking logic therefore exacerbates the sometimes already precarious position of freelance translators (Moorkens 2017) by following the model of control of salaried employment while still denying translators the greater security usually provided by in-house positions.

Although the problems of external tracking were more salient among freelance translators, distinguishing between external and internal tracking was also important for the in-house group. It was common practice at the company involved in our study for translators to use time trackers and discuss productivity in group and individual meetings. The interviews suggested that these discussions could unsurprisingly put translators under some degree of pressure. Translators drew a distinction, however, between monitoring their own productivity within the company and any practices that could in principle be controlled externally. The translation management system used in the company kept track of progress for each project, which translators regarded as standard. The time trackers complemented the information and allowed translators to record how much time they spent on any tasks not directly captured by progress reports generated by the management system. The in-house group did not have particularly negative views about these practices: "these productivity discussions help us […] to realise not only our productivity as [translators] or project managers but also if [the] client sends us projects that are profitable" (IH08). It is worth noting, however, how in-house translators were themselves in control of time trackers. Productivity measurement practices were also internal to the company, whose size of 10-15 employees, we posit, may help to give productivity measurement practices an internal ethos. On the other hand, like freelance translators, those who worked in-house were also largely against any productivity monitoring overseen by external stakeholders.

In summary, translators' comments in relation to internal vs external tracking show how the concept of productivity, the way it is measured, and what is done with the information are intricately connected to issues of trust (Alonso 2016; Abdallah & Koskinen 2007) where trust and external tracking are often in direct opposition.

4.3. *Invisible tasks*

Another topic frequently discussed in our focus-group interviews concerned the volume of ancillary tasks that translators must perform and that are not seen, strictly speaking, as part of the translation process. These usually included opening and preparing projects online or in a CAT tool, becoming familiarised with style guides or following other types of requirements. In all these cases, productivity tools risked disregarding the additional tasks translators had to perform. In a practical sense, this is because many of these tasks are outside the scope of the productivity measuring features of CAT tools, which can have consequences for translators as well as clients.

When translators wish to monitor their productivity and are in control of the process, they can work around any unlogged tasks by logging them manually—for example, with online timers, as mentioned in § 4.2. Ancillary and potentially unaccounted tasks are more than just a technical issue, however. In pricing models where translators' billable time consists solely of the time they spent on a client-controlled interface, this model may penalise translators in projects that require many administrative tasks that do not involve working on the text. Our focus groups suggest the number of such tasks is considerable, which is problematic given the usual emphasis on text length as a measure of work volumes. Indeed, translation technologies like translation memory and MT can obfuscate the relationship between word counts and the amount of work a task requires. This may in part explain the industry's increasing interest in temporal measures, which can be favourable to translators in cases where few words require high levels of effort. If translation time is monitored through external tracking, however, the benefits of temporal measurements are undermined if not cancelled altogether, not only by the way this can encroach upon professional freedom but also for the pragmatic reason that these measurements can be incomplete.

In some cases, even tasks that are more widely regarded as part of the translation process, such as research, may fail to be properly accounted for by activity tracking. Depending on how logging features are set up, they may stop tracking the time after a maximum period without any keyboard or mouse activity. In these cases, if translators are paid based on this measurement and go over the inactivity threshold while, for instance, checking reference materials, the time spent on research would be effectively invisible and thereby risks going unpaid. A freelance translator mentioned a specific case of this nature. One of their clients had moved from paying per word to a behavioural system where translators were paid based on their activity. The new tracking system failed to remunerate translators for intervals with no typing, even if the supposedly idle interval corresponded to the time translators dedicated to research. In this case, the behavioural system was a factor in why some translators dropped out of the team:

> [...] they say, "yeah, the tracker will only work when you type". So it doesn't take into account any research that you [do] [...] nobody is working for them anymore [...] So, it really depends [on] how knowledgeable the client is (FL05).

This example illustrates how weaknesses in the client's understanding of translation coupled with weaknesses in behavioural pricing systems may be detrimental to all those involved, including clients themselves.

4.4. *Conceptualising productivity*

Many of the issues posed by using behavioural data in translation revolve around what is understood by *productivity* and how this is operationalised. We were therefore interested in what *being productive* meant for translators. In our focus groups, we asked them what would prompt them, at the end of a working day, to think the day had been productive. Given the different circumstances of in-house and freelance translators, answers by these two groups unsurprisingly focused on different factors. In-house translators made several comments concerning the importance of considering not only their activity but the quality of their textual output. Despite attempts to standardise target quality expectations with labels such as "good enough" and "human quality" in relation to MT post-editing (Massardo et al 2016), in practice, client expectations often fall onto a spectrum (Vieira & Alonso 2018). This means that, to be methodologically effective, measurements of translation productivity need to account for the multitude of contexts that could influence expectations of quality. Furthermore, translation quality is not the only factor expected to influence productivity. Translation behaviour is

also likely to vary, for instance, according to the quality of reference materials and the amount, quality and timeliness of support provided by project managers. In principle, these factors are not impossible to measure, although their complexity undermines attempts to benchmark productivity on a large scale. This, in turn, risks imposing models of behaviour that do not necessarily fit all circumstances.

Indeed, translators mentioned the importance of making concessions to accommodate unexpected situations that would force them to work in ways that might deviate from what is considered "standard". This may be due to factors such as a particularly challenging source text or unusual client requirements or expectations. Translators' working patterns can also vary because of more general circumstances such as the Covid-19 pandemic—for example, because of increased caring responsibilities when schools were closed during national lockdowns. The list of factors that affect quantification of behaviour should also include personal circumstances, such as physical or mental disabilities, even though this was not reported in this study. In short, the increasing use of behavioural data to measure and predict productivity favours a numbers-oriented view of translation that can devalue its human factors:

> [...] when you try to reduce everything to just a number... you're missing a lot of stuff, you know, and... the translator... that profession... is a very human one and you have to take into account the human factors because, you know, it's humans doing the work. (IH01)

Translators' understanding of *productivity* in our study challenged simplistic, client-oriented models. Profitability was at the forefront of freelancers' minds. One translator in this group mentioned: "I think the concept is strictly connected to the amount of money you earn compared to the number of hours you work" (FL11). Some translators thought that activity tracking was useful precisely because they could monitor their own profitability as freelancers, which in some cases came as a positive surprise: "for me it was really great to know that 'oh, I'm actually making a decent living and, you know, for this kind of project I can afford to only work, like, three hours a day'" (FL02). Comments of this nature have two important implications. First, they point to reaping the benefits of one's own productivity as a critical factor in "being productive". Second, they illustrate a key aspect of productivity, namely that it is relative.

Productivity can be framed differently depending on the position different stakeholders occupy in production networks. Assuming the same subject matter, text type and use of technologies, someone who translates 350 words an hour, for instance, would in many cases be considered "more productive" than someone who translates 200. From a single client's perspective, assuming equal levels of target-text quality and equal hourly pay, the translator who delivers 350 words an hour would be preferred. If the translator who delivers 200 words/hour commands a higher rate of pay, again assuming equal quality this translator would be even less attractive financially from the client's perspective. Conversely, the 200 words/hour translator would in fact be the more productive of the two in this case, if translator profitability is considered. For freelance translators, therefore, productivity is less about how many words they can translate per unit of time and more about the financial return they have on that time or those words. When productivity is tracked externally, this tracking will rarely have translators' individual profitability as its driver, however. If an algorithm raises translators' hourly rate *subject to* higher words/time throughput (see § 1), this means that throughput rather than translator profitability is the key factor in the equation. This is not necessarily malicious, and indeed the underlying mechanisms of this logic are not new. This pricing model simply reflects the market paradigm of efficiency and profit maximisation. What is relatively new, however, is the way behavioural data can propagate this logic at scale, often under the presumed objectivity of data analytics, and in ways that can be disproportionally detrimental to freelance translators.

5. Discussion

When translators self-track without unwelcome external interference, they may be able to gain greater insight into what makes them more productive or their work more profitable. These can be classed as translators' own motivations to track behaviour, particularly if they are freelancers. For them, internal tracking may constitute part of the usual principles of running a business which, by the same token, can apply to in-house translators at a company level. When translators do not set the

principles and motivations for this tracking, however, they in part cede control over how they work. Pragmatically, this may happen, for example, because translators need to be mindful of any tasks that might go untracked and, depending on the pricing model, unpaid. This may also happen because translators need to engage in artefact behaviours to satisfy arbitrary expectations. Ceding control of working practices therefore leaves translators exposed to the methodological limitations of activity tracking.

As mentioned in § 4.3, part of the problem of using tracking methods for purposes such as calculating pay is that the method may miss important aspects of a job, thereby potentially failing to represent the effort translators invest in it. When used for algorithmic decision making or productivity benchmarking, behavioural methods may also be limited by the complexities of accounting for the wide range of factors that are expected to influence the translation process. External tracking therefore involves a methodological conundrum. To be accurate, the tracking would need to be sensitive to different circumstances and capture with precision any tasks that take place outside the translation tool. However, this would exacerbate some of the method's already negative consequences pointed out by translators by making it even more invasive and therefore undesirable.

Seen in the wider context of how professional translation practice has evolved, the negative implications of tracking methods reflect important characteristics of how translation services are currently structured, which may also apply to other industries that can operate online or which now involve platform-mediated work (Srnicek 2017). Before cloud and server technologies, translators largely had to acquire their own CAT tools and install them locally on their computer. This may in many cases represent an obstacle. The financial investment can be considerable, especially for a fledgling translator. Client-owned cloud and server technologies at first sight solve this problem by allowing translators to access the technology without having to pay for it or at least with reduced upfront costs. If this means being constantly monitored online, however, renouncing control of the technology also incurs a price since it can at least partly mean renouncing control of behaviour. Furthermore, this is to do not only with isolated translation tasks but also with long-term changes in the collective paradigm that governs what is expected of translators. Data from a single translator is not of much use to set productivity benchmarks or to train predictive models that can profile translators based on speed. Like in certain conceptualisations of autonomy where the focus is on collective self-governance (Flikschuh 2012), the use of activity data in translation for benchmarking and other external purposes transcends the individual data provider. It affects all those who in the future might be subject to models that dictate what constitutes standard behaviour or what level of speed is worthy of higher pay. We therefore call for greater awareness of how structural imbalances of power in the language industry can turn behavioural data into an instrument of control.

These power imbalances are unsurprisingly connected to imbalances of knowledge. Even with the growing importance of translation technologies, it is arguably not realistic to expect freelance translators to have the programming skills to develop their own tools from scratch. This means that even by involving translators in technological development, it is difficult for them to control all aspects of the technology or its motivation to exist. The age-old interaction between knowledge and power therefore also exposes the nuanced nature of the cloud, which can at once be controlling but also bridge gaps in knowledge and hardware infrastructure to put some level of technological implementation in the hands of translators, for example in relation to MT training frameworks (see Kenny & Doherty 2014). In all these cases, it is necessary for different stakeholders to consider the long-term effects of new methods and technologies and to keep them under review, which applies to the private sector as much as it does to academic research.

6. Conclusion

This article reports on an empirical study that aimed to investigate how translation activity tracking could itself change work behaviour, and how it might have wider implications for translators.

In relation to our first aim, we highlight a series of pragmatic impacts that tracking can have on translators' working methods. Translators' empirically informed comments in our focus-group interviews largely reflected organisational behaviour research. Some translators reported trying to "gamify" their own productivity by trying to use the activity tracking feedback as a way of improving their speed. However, based on translators' own assessment of the real-life experience of using

tracking tools during the study, gamification had inconsistent and short-lived effects, not least because it could be detrimental to target-text quality. In some cases, tracking the translation process was also found to be demotivating.

Regarding our second aim, we call attention to a series of wide-ranging consequences of using behavioural data in professional translation. We argue that attempts to model professional behaviour for purposes such as pricing and benchmarking can undermine translator autonomy by propagating the commissioners' paradigm of productivity while installing new collective logics that dictate what is expected of translators and what can be regarded as desirable behaviour in the future.

Although, as mentioned in § 2, we draw a distinction between uses of behavioural data in the industry and in translation process research, some of the results presented here also have implications for how previous academic findings are used and how future studies are conducted, including work by the authors. There is room for future research to look at how translation tools and working practices can be more inclusive and at how potential misuses of technologies like the cloud can be avoided, for example by updating existing regulations and designing new ones that more explicitly account for the digital milieu in which commercial translation occurs. There is also room for translation research to raise further awareness of the potential consequences of exploiting behavioural data in ways that favour perspectives of productivity that can be detrimental to translators.

Many negative consequences of translation activity tracking have been covered in this article together with positive ones that are subject to the context in which the data is collected. Importantly, the negative consequences of behaviour quantification can affect all those who depend on translation professionals to have the freedom to exert their expertise to the best of their ability to add value to services and products. This is not therefore just a matter for translators. Ultimately, clients also have a stake in translator autonomy.

References
Abdallah, Kristiina, and Kaisa Koskinen. 2007. "Managing Trust: Translating and the network economy." *Meta* 52 (4): 673–687. https://doi.org/https://doi.org/10.7202/017692ar.

Adams, William C. 2015. "Conducting Semi-Structured Interviews." In *Handbook of Practical Program Evaluation.* Edited by K. E. Newcomer, H. P. Hatry, and J. S. Wholey, 492–505. Hoboken, NJ: Wiley. https://doi.org/10.1002/9781119171386.ch19

Alonso, Elisa. 2016. "Conflict, Opacity and Mistrust in the Digital Management of Professional Translation Projects." *Translation & Interpreting* 8 (1): 19–29. https://doi.org/10.12807/ti.108201.2016.a02.

Ashton, Kevin. 2009. "That 'Internet of Things' Thing: In the real world, things matter more than ideas." *RFID Journal*. Accessed 16 September 2020. https://www.rfidjournal.com/that-internet-of-things-thing.

Baclawski, Kenneth. 2018. "The Observer Effect." In *Proceedings of the 2018 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA)*, 11–14 June 2018, Boston, MA. Edited by G. L. Rogova, C. Lebiere, O. E. Gundersen, A. Salfinger, and K. Baclawski, 83–89. Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/COGSIMA.2018.8423983.

Bartlett, Benjamin. 2020. "How do Hourly Rates Work? How can i increase mine?". *Unbabel Support*. Accessed 06 November 2020. https://help.unbabel.com/hc/en-us/articles/360003294194.

Bershidsky, Leonid. 2019. "A European Court Ruling that Requires Companies to Track Employees' Exact Hours Could Yield Useful Productivity Data." *Bloomberg*. Accessed 22 December 2020. https://www.bloomberg.com/opinion/articles/2019-05-16/eu-time-tracking-ruling-could-generate-useful-productivity-data.

Bevir, Mark. 1999. "Foucault and Critique: Deploying agency against autonomy." *Political Theory* 27 (1): 65–84. http://www.jstor.org/stable/192161.

Buzelin, Hélène. 2011. "Agents of Translation." In *Handbook of Translation Studies*, Volume 2. Edited by Y. Gambier and L. Van Doorslaer, 6–13. Amsterdam: John Benjamins.

Christman, John. 2020. "Autonomy in Moral and Political Philosophy." In *Stanford Encyclopedia of Philosophy.* Edited by E. N. Zalta. Accessed 09 April 2021. https://plato.stanford.edu/archives/fall2020/entries/autonomy-moral/.

do Carmo, Félix. 2020. "'Time is money' and the value of translation." *Translation Spaces* 9 (1): 35–57. https://doi.org/10.1075/ts.00020.car

EC, CIOL, and ITI. 2017. *2016 UK Translator Survey: Final report.* European Commission Representation in the UK, the Chartered Institute of Linguists & Institue of Translation and Interpreting. Accessed 09 April 2021. http://www.ciol.org.uk/sites/default/files/UKTS2016-Final-Report-Web.pdf.

Flikschuh, Katrin. 2012. "Personal Autonomy and Public Authority." In *Kant on Moral Autonomy.* Edited by O. Sensen, 169–190. Cambridge: Cambridge University Press.

Foucault, Michel. 1975. *Surveiller et punir: naissance de la prison* [Discipline and Punish: The birth of the prison]. Paris: Gallimard.

Franklin, Stan, and Art Graesser. 1997. "Is it an Agent, or Just a Program? A taxonomy for autonomous agents." In *Intelligent Agents III Agent Theories, Architectures, and Languages.* Edited by J. P. Müller, M. J. Wooldridge, and N. R. Jennings, 21–35. Berlin: Springer.

García, Ignacio. 2017. "Translating in the Cloud Age: Online marketplaces." *HERMES Journal of Language and Communication in Business* 56: 59–70. https://doi.org/10.7146/hjlcb.v0i56.97202

Hansen-Schirra, Silvia. 2017. "EEG and Universal Language Processing in Translation." In *The Handbook of Translation and Cognition.* Edited by J. W. Schwieter and A. Ferreira, 232–247. Hoboken, NJ: Wiley.

Huang, Ronggui. 2016. *RQDA: R-Based Qualitative Data Analysis.* R Package Version 0.2-8. http://rqda.r-forge.r-project.org/

Kenny, Dorothy, and Stephen Doherty. 2014. "Statistical Machine Translation in the Translation Curriculum: Overcoming obstacles and empowering translators." *The Interpreter and Translator Trainer* 8 (2): 276–294. https://doi.org/10.1080/1750399X.2014.936112.

Kinnunen, Tuija, and Kaisa Koskinen, eds. 2010. *Translators' Agency*. Tampere: Tampere University Press.

Labov, William. 1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

LeBlanc, Matthieu. 2017. "'I Can't Get No Satisfaction!' — Should we blame translation technologies or shifting business practices?" In *Human Issues in Translation Technology.* Edited by D. Kenny, 45–62. London: Routledge.

Luck, Michael, and Mark d'Inverno. 1995. "A Formal Framework for Agency and Autonomy." In *Proceedings of the First International Conference on Multi-Agent Systems*, 254–260. Cambridge, MA: Association for the Advancement of Artificial Intelligence (AAAI). Accessed 09 April 2021. https://www.aaai.org/Papers/ICMAS/1995/ICMAS95-034.pdf.

Massardo, Isabella, Jaap van der Meer, Sharon O'Brien, Fred Hollowood, Nora Aranberri, and Katrin Drescher. 2016. *MT Post-Editing Guidelines.* Amsterdam: TAUS Signature Editions.

memoQ. n.d. "Create Editing Time Report." *memoQ*. Accessed 12 November 2020. https://docs.memoq.com/current/en/Places/create-editing-time-report.html.

Moorkens, Joss. 2017. "Under Pressure: Translation in times of austerity." *Perspectives* 25 (3): 464–477. https://doi.org/10.1080/0907676X.2017.1285331.

Moorkens, Joss. 2020. "A Tiny Cog in a Large Machine": Digital Taylorism in the translation industry." *Translation Spaces* 9 (1): 12–34. https://doi.org/10.1075/ts.00019.moo.

Moran, John. 2018. "Extensive CAT Tool Logging: Big Brother or language technology evaluation panacea?" In *ITI Research Network e-book 2018: The human and the machine*, 3. Institute of Translation and Interpreting. Accessed 09 April 2021. https://www.iti.org.uk/resource/iti-research-ebook-translation-interpreting-2018.html.

O'Brien, Sharon. 2011. "Towards Predicting Post-Editing Productivity." *Machine Translation* 25 (3): 197–215. https://doi.org/10.1007/s10590-011-9096-7.

Olohan, Maeve. 2011. "Translators and Translation Technology: The dance of agency." *Translation Studies* 4 (3): 342–357. https://doi.org/10.1080/14781700.2011.589656.

O'Neil, Cathy. 2016. *Weapons of Math Destruction: How big data increases inequality and threatens democracy*. New York: Crown.

Ragni, Valentina, Elisa Alonso, and Lucas Nunes Vieira. *Translation Productivity: Tools, tracking, ethics*. Forthcoming.

Ranganathan, Aruna, and Alan Benson. 2020. "A Numbers Game: Quantification of work, auto-gamification, and worker productivity." *American Sociological Review* 85 (4): 573–609. https://doi.org/10.1177/0003122420936665.

Salkind, Neil J. 2010. *Encyclopedia of Research Design* (Vols. 1-0). Thousand Oaks, CA: SAGE Publications. doi:10.4135/9781412961288

RWS. n.d. *Translation Productivity*. RWS. Accessed 21 June. https://community.sdl.com/product-groups/translationproductivity/.

RWS Community. n.d. *Qualitivity*. RWS. Accessed 12 November 2020. https://community.sdl.com/product-groups/translationproductivity/w/customer-experience/2251/qualitivity.

Srnicek, Nick. 2017. *Platform Capitalism*. Cambridge: Polity.

Starkmann, Angela. 2020. "Need a Productivity Boost? Read how memoQ's main features can help." *memoQ Blog*. Accessed 09 April 2021. https://blog.memoq.com/productivity-boost-memoq-features#:~:text=memoQ%20is%20a%20great%20CAT,learning%20how%20to%20use%20it.

Straker. n.d. *XRAY - Data analytics engine*. Straker. Accessed 16 September 2020. https://www.strakertranslations.com/ray/xray/.

Swaine, Lucas. 2016. "The Origins of Autonomy." *History of Political Thought* 37 (2): 216–237.

TAUS. 2020. *Real-time Translation Analytics*. Translation Automation User Society. Accessed 09 April 2021. https://www.taus.net/data/dqf.

Toral, Antonio, Martijn Wieling, and Andy Way. 2018. "Post-editing Effort of a Novel With Statistical and Neural Machine Translation." *Frontiers in Digital Humanities* 5: 9. Accessed April 09, 2021. doi:10.3389/fdigh.2018.00009.

Vieira, Lucas Nunes, and Elisa Alonso. 2018. *The Use of Machine Translation in Human Translation Workflows: Practices, perceptions and knowledge exchange.* Institute of Translation and Interpreting. Accessed 09 April 2021. https://www.iti.org.uk/resource/the-use-of-machine-translation-in-human-translation-workflows.html.

Webber, Ashleigh. 2020. "PwC Facial Recognition Tool Criticised for Home Working Privacy Invasion." *Personnel Today*. DVV Media International. Accessed 09 April 2021. https://www.personneltoday.com/hr/pwc-facial-recognition-tool-criticised-for-home-working-privacy-invasion/.

Zheng, Binghan, Sandra Báez, Li Su, Xia Xiang, Susanne Weis, Agustín Ibáñez, and Adolfo M. García. 2020. "Semantic and Attentional Networks in Bilingual Processing: fMRI connectivity signatures of translation directionality." *Brain and Cognition* 143: 105584. https://doi.org/10.1016/j.bandc.2020.105584.

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The fight for a human future at the new frontier of power*. London: Profile Books.

Authors' addresses

Lucas Nunes Vieira
University of Bristol
School of Modern Languages
17 Woodland Road, BS8 1TE

l.nunesvieira@bristol.ac.uk

Valentina Ragni
University of Bristol
School of Modern Languages

17 Woodland Road, BS8 1TE

v.ragni@bristol.ac.uk

Elisa Alonso
Universidad Pablo de Olavide
Ctra. de Utrera, km 1
41013 Seville, SPAIN

elialonso@upo.es

Biographical notes

Lucas Nunes Vieira is a Senior Lecturer in Translation Studies in the School of Modern Languages at the University of Bristol, United Kingdom.

Valentina Ragni is a Research Associate in the School of Modern Languages at the University of Bristol, United Kingdom.

Elisa Alonso is a Senior Lecturer in Translation Studies in the Department of Philology and Translation at Universidad Pablo de Olavide, Spain.