Virtual Humans as Co-Workers: A Novel Methodology to Study Peer Effects

Özgür Gürerk* Andrea Bönsch† Thomas Kittsteiner‡ Andreas Staffeldt‡

October 24, 2018

Abstract

We introduce a novel methodology to study peer effects. Using virtual reality technology, we create a naturalistic work setting in an immersive virtual environment where we embed a computer-generated virtual human as the co-worker of a human subject, both performing a sorting task at a conveyor belt. In our setup, subjects observe the virtual peer, while the virtual human is not observing them. In two treatments, human subjects observe either a low productive or a high productive virtual peer. We find that human subjects rate their presence feeling of "being there" in the immersive virtual environment as natural. Subjects also recognize that virtual peers in our two treatments showed different productivities. We do not find a general treatment effect on productivity. However, we find that competitive subjects display higher performance when they are in the presence of a highly productive peer - compared to when they observe a low productive peer. We use tracking data to learn about the subjects' body movements. Analyzing hand and head data, we show that competitive subjects are more careful in the sorting task than non-competitive subjects. We also discuss some VR related methodological issues.

Keywords: peer effects, real effort, virtual reality, virtual human, reflection problem, immersive virtual environment

JEL Codes: C91, J24, M50

Acknowledgments: We thank Jonathan Wendt for his assistance in conducting the experiment, and Lucas Braun for preparing the raw data for the analysis. We also thank Almut Balleer, Robert Böhm, Bernd Irlenbusch, the participants of the seminars in Bamberg, Erfurt, Göttingen, Karlsruhe; the conference participants at the 2015 Economic Science Association Meetings in Sydney, Heidelberg, and Dallas; and the participants of the 15th TIBER Symposium in Tilburg, for their valuable comments. We acknowledge the financial support from the "Excellence Initiative (ZUK II)" of the German Science Foundation (DFG).

*Corresponding author, RWTH Aachen University, School of Business and Economics, Aachen, Germany. E-mail: ozgur.gurerk@rwth-aachen.de

†RWTH Aachen University, Visual Computing Institute, and JARA-HPC, Aachen, Germany.

‡RWTH Aachen University, School of Business and Economics, Aachen, Germany.

1 Introduction

Peer effects are observed in a plethora of domains, such as financial decision making (Bursztyn, Ederer, Ferman, & Yuchtman, 2014), academic success of students (Zimmerman, 2003), or in health-related choices (Trogdon, Nonnemaker, & Pais, 2008). One important domain where peer effects can occur is the workplace. Workers can influence each other in their work performance when they receive information about or directly observe the respective peer's performance. How strong is the influence of a "rotten apple" on her peers' performances? Does a top performer induce fellow workers to high effort levels? Understanding the size and nature of such peer effects allows for a more effective organization of labor by answering important questions, such as whether shifts should be composed of workers with homogeneous or heterogeneous productivities/potentials.

We analyze peer effects in an organizational framework where workers work, side by side, on identical tasks but do not interact in any way, except that one of the (two) workers can observe the other worker's activity. In particular, workers neither complement nor substitute each other in the production process, and receive a fixed and performance independent wage. Spillovers that may occur in such situations are referred to as "pure" peer effects (Charness & Kuhn, 2011; Falk & Ichino, 2006).

In this study, we introduce a new experimental methodology for testing (pure) peer effects. We create a naturalistic work setting in an *immersive virtual environment* (abbrev. as IVE), where human subjects work together with peers who are computer-generated *virtual humans* (abbrev. as VH). Subjects perform a real-effort sorting task with their dominant hand, while at the same time the experimenter has control over the actions of the VH via an algorithm. This control over the VH eliminates the reflection problem (Manski, 1993): while the VH may influence the human subject, the virtual peer cannot be influenced by the human subject.[1]

Specifically, we address the question whether and to what extent a subject's work performance is affected by a peer whose performance can be observed, but where the subject's performance is unobserved by the peer. To control for heterogeneity in the performance of the specific task at hand, we design an experiment consisting of two phases. In the first phase (the initial phase), subjects perform the sorting task alone. In the second phase (the peer phase), they perform the task in the presence of a peer, working independently from each other, and a worker's payoff is fixed. In two treatments, we model the subject's virtual peer either as a low productive or a highly productive worker. We refer to these treatments as the SLOW and the FAST treatment, respectively.

Based on a post-experimental survey, we find that most subjects perceive the virtual

[1] In Section B in the Appendix, we discuss in more detail how the experimental literature approached the reflection problem, and how we handle it.

peer as human-like and the IVE as naturalistic. Subjects evaluate the VH they observe in the FAST treatment indeed as the more productive one – compared to the VH they watch in the SLOW treatment. This means subjects perceive the difference in the performances of both virtual peers exactly as intended. Furthermore, a significant share of subjects, to which we refer as the "competitive" subjects, reports that they wanted to be more successful than the virtual peer , which we interpret as further evidence that subjects acknowledge and react to the presence of VHs.

Over all, we find no statistically significant peer effect between treatments. Interestingly, competitive subjects display significantly higher work performance in the presence of a high productive virtual peer, compared to when they work in the presence of a low productive peer.

Our methodology allows us to collect tracking data in the IVE, in particular, the position and orientation of a subject's head and hand. These data allow us to define novel performance measures and refine our findings. Using tracking data, we find that competitive subjects display more careful sorting behavior than non-competitive subjects.

We make two important contributions to the literature. First, we introduce a novel experimental methodology to investigate peer effects in a naturalistic setting, but without losing any experimental control. We refine our findings using the tracking possibilities of VR technology. Second, we contribute to the literature on peer effects, by reporting a peer effect in competitive subjects.

2 Related literature on peer effects

Peer effects have been attributed to *peer pressure* which may cause a positive productivity effect, due to the workers' general dislike of providing lower effort than the peer(s).[2] Peer pressure, however, may also induce negative productivity effects, in particular, due to discouragement (Georganas, Tonin, & Vlassopoulos, 2015). Peer effects have also been attributed to *knowledge spillovers* causing a positive peer effect (Azoulay, Graff Zivin, & Wang, 2010; Jackson & Bruegmann, 2009), and to *free riding* causing a negative peer effect (Mas & Moretti, 2009).

Empirical evidence on the existence and the nature of peer effects is mixed. In their meta-study, Herbst and Mas (2015) include 17 peer effects reporting positive significant peer effects, while 16 studies do not find any peer influence, one study reporting negative significant peer effects. Herbst and Mas (2015) find that positive peer effects are particularly present in studies with group piece rate and fixed wage produce. In a data set comprising workers from different occupations, firms and sectors Cornelissen, Dustmann,

[2]This dislike can stem from the threat of social sanctions (Kandel & Lazear, 1992).

and Schönberg (2017) show that positive peer effects are more pronounced in low-skilled occupations for which the authors argue that peer pressure is more relevant than for high-skilled occupations.

In our experiment, we focus on a specific kind of peer effect caused by a worker's dislike to exert little effort if an observed peer provides high effort. We do not investigate possible psychological origins of such a dislike but are solely interested in its presence (or absence) in a highly controlled environment that eliminates other causes by design. The previous evidence for such a dislike to provide low effort if the observed peer does not is mixed. In a field study about peer effects among cashiers of a grocery store, Mas and Moretti (2009) do not find evidence for such a positive effect and conclude that prosocial behavior is not supported as a possible cause. This latter finding is, at least partially, in contrast to previous lab experiments. Georganas et al. (2015) report evidence that subjects who perform a real effort task, are paid individual piece-rates, and observe someone else performing the same task (but are not observed themselves) decrease own productivity when seeing low-productive peers, and increase own productivity when observing high-performers.[3] Also, Gerhards and Gravert (2017) show that peer effects on the side of the observing peer may occur. In their experimental setup, workers who observe other workers skip fewer single instances of a task (solving anagrams) and go with the next one, than those in the control treatment where workers perform the tasks alone. In another real-effort lab experiment, Van Veldhuizen, Oosterbeek, and Sonnemans (2018) find no evidence that observing subjects positively react to observed subjects' performance.

A number of studies provide evidence for pure peer effects. Falk and Ichino (2006) conduct an experiment, in which subjects individually perform a real effort task for which they are each paid a fixed wage. They show that peers who work in sight of each other perform similarly, and they on average exert higher individual efforts than agents working alone. Kaur, Kremer, and Mullainathan (2010) report positive (pure) peer effects among data entry workers sitting next to each other, and (Bandiera, Barankay, & Rasul, 2010) find that fruit pickers' performances are positively correlated to their friends' performances on the same field.

The idea to investigate peer effects in IVEs was first proposed in a perspectives paper by Gürerk et al. (2014). The research agenda was briefly explained in Bönsch et al. (2017) reporting the basic idea and the first results of two different studies using virtual experiments, one of which is the present study. While we focus on pure peer effects in a fixed wage setting, without any payoff interdependency between workers, Graff, Grund, and Harbring (2018) study peer effects with piece-rate incentives, when workers perform alone,

[3] Different from our setting, observing subjects are explicitly informed about the productivity of their peer. In addition, the observer works after observing the peer, i.e.; workers do not act simultaneously as in our setting.

and in tournaments between a human subject and a VH. Different from our study, in Graff et al. (2018) workers are informed about own and the virtual peer's productivity in real time. Again, different from our setup, the virtual peer's productivity is adjusted to the base productivity of the human worker. The main similarity between their study and ours is that Graff et al. (2018) use a modified version of the IVE that was originally developed for the present paper.

3 Hypotheses

We investigate whether and how a worker's performance is influenced by the physical presence of another worker, referred to as co-worker/peer, whom the worker can observe performing an identical but otherwise unrelated task. In particular, we ask whether the observable productivity of a peer affects a worker's own work performance in a setup without any interdependency between co-worker and worker.

3.1 Standard economic prediction

A payoff maximizing worker does not react to a peer's observed performance as payments and work outcome of the worker and the peer are independent, which also excludes free-riding as a source of peer effects. Because of the fixed wage, there is no incentive to exert any effort at all. Ignoring dynamic aspects such as the possibility of learning during the course of the experiment, standard economic theory leads to the following hypothesis:

Hypothesis 1 *There are no differences in workers' average performances whether the subject is working in the presence of a low productive peer (in the SLOW treatment) or in the presence of a high productive peer (in the FAST treatment).*

3.2 Learning by imitation

In our setting, the worker can (implicitly) learn how to sort cubes by observing and imitating the virtual peer, though this is partly hampered by the fact that subjects were not able to fully observe their peer's actions as these were partly covered by the peer's body. Whereas subjects in our experiment cannot learn a more efficient sorting technique in the FAST treatment than in the SLOW treatment, they can observe more (but not better) sorting incidents. Based on this argument a better performance in the FAST treatment can be expected.

Hypothesis 2 *Workers performances are higher in the FAST treatment than in the SLOW treatment.*

3.3 Competitive motivations and the desire to win

In our setting, subjects do not receive feedback about their own and their peer's performances, and there is no extrinsic motivation to compete. However, some subjects may perceive the situation as if it was a head-to-head competition, and may be intrinsically motivated to "beat" the virtual peer. Agents who experience such a "desire to win" (Malhotra, 2010) may engage in costly effort activities, in some cases even irrationally, such as overbidding in auctions (Ku, Malhotra, & Murnighan, 2005).

To learn about their intrinsic motivations, in the questionnaire, we asked subjects whether they wanted to be more successful than the virtual human. We refer to subjects, who answered this question with yes as competitive subjects. We argue that competitive subjects may display a higher performance than subjects who are classified as non-competitive. As subjects need to perform better to outperform a high productive peer as compared to a low productive peer, we also expect competitive subjects to perform better in the FAST treatment than in the SLOW treatment.

Hypothesis 3

a) Competitive subjects perform better than other subjects.

b) Competitive subjects perform better in the FAST treatment than in the SLOW treatment.

4 History and methodology of VR experiments

In this section, we provide a brief history of experiments conducted in IVEs, and some introductory methodological remarks. Following some other scholars, we also refer to them as *VR experiments.* Behavioral research using VR experiments started in the 1990s in domains such as visual perception, spatial cognition, psychotherapy, education, and learning. For a comprehensive overview of early VR experiments in the 2000s and before, we refer to the book by Blascovich and Bailenson (2011).

4.1 VR experiments in social psychology

The use of IVEs in social psychology has been proposed by, e.g., Biocca and Levy (1995), and one of the first experiments was conducted by Loomis, Blascovich, and Beall (1999). But, why conduct VR experiments? Blascovich et al. (2002) consider IVEs as a tool that can overcome some of the long-term methodological problems of experimental (social psychology) behavioral research. First, IVEs can increase mundane realism without losing experimental control, i.e., in VR one can create experimental environments that more closely resemble reality while keeping the advantages of a controlled lab setting. Second, for studies involving human social interaction, in IVEs, one can create VHs, who act exactly in the same way in each observation, immune to be influenced by the participant. IVEs enable researchers to conduct experiments that otherwise would be too costly, too risky/dangerous (crossing a plank seemingly 30 feet over the ground), or simply impossible (flying over a scene). There is evidence, that subjects respond realistically when experiencing a plausible, safe, and controllable VR scenario (Slater et al., 2013).

4.2 VR experiments in economics

In economics, the idea of conducting experiments in IVEs has been proposed by Fiore, Harrison, Hughes, and Rutström (2009). They state the objective of *virtual experiments* as to "bridge the gap between the artefactual controls of laboratory experiments and the naturalistic domain of field experiments or direct field studies". This paper presents several arguments in favor of the use of IVEs in (experimental) economic research. The authors find that subjects can evaluate risks associated with wildfires more accurately when they can watch realistic and physically accurate models of wildfires presented on 3D screens, compared to when only still pictures are provided.[4]

A variant of the basic design of the present study (without co-workers) is used to investigate a question in the context of behavioral operations management (DeHoratius, Gürerk, Honhon, & Hyndman, 2018).

The first review of virtual experiments in economics, recently published by Innocenti (2017) nicely discusses previous experiments conducted in "low" IVEs like *virtual worlds* and using 3D monitors, as well as experiments conducted in "high" IVEs utilizing CAVEs and HMDs. Experiments conducted in virtual worlds differ from experiments conducted in IVEs, as in virtual-worlds studies, subjects usually sit in front of a 2D monitor and navigate by using the keyboard or the mouse. For the research potential of virtual worlds, see Bainbridge (2007). Harrison, Haruvy, and Rutström (2011) provide a methodological remark on the potential use of VR in behavioral research, discussing virtual world experiments, too.

4.3 Display devices

An IVE perceptionally surrounds the subject such that the person experiencing it has a strong feeling of "being there" in the synthetically created situation. There are two main

[4]Our setup is different from that in Fiore et al. (2009). First, our virtual environment offers a higher degree of immersion since subjects are not looking on small flat 3D monitors, but they enter a VR chamber, surrounded by large 3D projection walls. Second, our subjects are not only visually stimulated as in Fiore et al. (2009), but they can move in the CAVE and naturally interact with virtual objects, by using their hands (which facilitates naturalistic real-effort tasks). Third, our subjects are confronted with virtual humans.

methods to experience IVEs, either entering a CAVE[5], i.e., a VR chamber; or using a head-mounted display (abbrev. HMD).

One major difference between the two systems is that the CAVE can usually provide the correct visual perspective only for just one person.[6] Even though some CAVEs can accommodate for more than one person at the same time, only one user will have the correct perspective whereas the other users will enjoy some blurred vision, the less so, the closer they stand to the person whose sight has the correct perspective. Different from CAVEs, it is relatively easy to connect several HMDs. The connectivity of many HMDs allows running experiments where many subjects meet in the same virtual place though they physically may be in different places. Collaboration experiments are possible where several subjects interact with objects and with each other.

Another difference between CAVEs and HMDs concerns the costs. CAVEs that are relatively large, and with relatively high display resolutions can be quite expensive, as the system at the RWTH Aachen University, which costs were more than 1 Million Dollars as it was installed in 2012.[7]. CAVEs of smaller size are more affordable, such as the CAVE of the DAF Technology Lab at the University of Tilburg. Running costs of a CAVE depend on the technology of the projectors. The running costs of our experiment, which ran eight days from 10 to 18 o'clock, were estimated with over 1000 Euros. Running an experiment in a CAVE is also associated with some personnel costs. In our case, two research assistants, one instructor, one for the CAVE were busy for eight full days.

Recently, HMDs became very affordable. The prices of the *HTC Vive* system or the *Oculus Rift* system are well below 1000 US Dollars. Using such HMDs as seen in Figure 1, researchers can conduct VR experiments at a low cost. Running costs of HMDs are also lower than the running costs of CAVE. First, there are no bulbs that have a restricted life time. Second, one can run several HMDs in parallel, which can decrease the personnel costs per session. Authors of this study, as well as other behavioral scientists use HMDs in their research, (see, e.g., Gürerk and Kasulke (2017)) such as the *EVENT lab* in Barcelona, the *Virtual Embodiment Lab* at Cornell University, the *Virtual Human Interaction Lab* at Stanford University, and the *Behavioral Lab* in St. Gallen.

Our experiment can also be run using an HMD which enables reproducibility of our study not only in a CAVE. In a study which one of the authors prepare, we will run an experiment similar to introduced in this study with an HMD. In particular, for a Vive System, a slightly different version of the sorting task is already available as ready-to-run

[5]CAVE is an acronym for Cave Automatic Virtual Environment, see Cruz-Neira, Sandin, and DeFanti (1993) for the basic working principles of a CAVE

[6]As the projector technology is quickly evolving, dual user CAVEs may emerge soon.

[7]We will elaborate more on the technological details of this specific CAVE in Section 5.1

Figure 1: A person using an HMD

version upon request.[8]

4.4 Immersion and presence

Two important concepts define the validity of a virtual environment: *immersion* and *presence*. Immersion relates to objectively measurable, technical aspects of the virtual environment, for example, the resolution of the displays, whereas presence describes the intensity of the subjective experience, i.e., the participants' subjective feeling of actually "being there". Given the same IVE, i.e., for a given degree of immersion, subjects may experience presence differently (Slater & Wilbur, 1997). A recent review by Cummings and Bailenson (2015) covering 83 VR studies reports that some aspects of immersion such as the accuracy of participant-tracking, or the field of view of the visual displays have a greater impact on presence than other aspects, such as the quality of the visual and auditory content.

4.5 On virtual humans

Our study is based on the assumption that subjects' actions in IVEs and their interactions with VHs are informative about how subjects behave in reality. Whereas we investigate whether and to what extent participants in our experiment consider the specific IVE as "natural", we do not validate the method as such. Here, we rely on an already significant body of research showing that IVEs generally are perceived as realistic and that subjects

[8]Watch here a clip showing a person with an HMD who performs a similar sorting task as in our experiment: https://www.youtube.com/watch?v=VlNxx33ZLFU.

consider and react to VHs in a similar way as to humans, see, e.g., the comprehensive review by Slater and Sanchez-Vives (2016). Our findings on the validity of the IVE and the acceptance of VHs by participants are reported in Section 6.1.

The term *personification* describes the graphical/visual quality of a virtual human, and how natural it appears. The quality of personification influences the acceptance of virtual humans by their human counterparts and how humans interact with them (Kasap & Magnenat-Thalmann, 2007). To be considered human-like, virtual humans should not only look like real humans but also act/move human-like. Therefore, plausible facial expressions, correct eye movements and natural movements of the body are important for the quality of personification. For our experiments, facial expressions do not matter as the virtual peer is visible only from behind whereas natural body movement is crucial achieving a sufficiently high quality of personification.[9]

Previous studies show that participants accept virtual humans as interaction partners, and treat them similar to humans. Bailenson, Blascovich, Beall, and Loomis (2003), for example, measure the interpersonal distance between the human participant and a VH, when the participant should walk around the VH in a virtual room. As is the case with human-to-human studies, participants move closer to the VH when the VH is not having any eye contact with them – as compared to a situation where the VH is looking at them.

In general, when participants encounter virtual humans, one can observe an increase in arousal, which is the physiological and psychological state of being awake, and an indicator of human emotions. Slater et al. (2006), for example, measure the increase in arousal through physiological responses such as skin conductance and heart rate variability. For a comprehensive review of the evidence on humans' perception of virtual humans in social interactions, see Bombari, Schmid Mast, Canadas, and Bachmann (2015).

5 Experimental setup and procedures

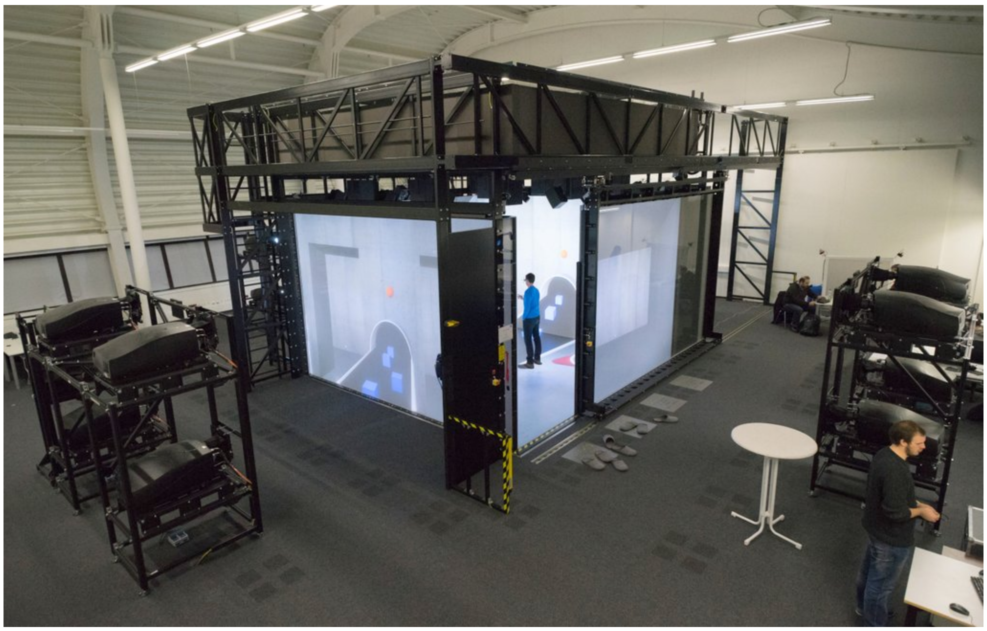
5.1 The virtual environment

We conduct our experiment in the surround-projection room *aixCAVE* at the RWTH Aachen University (see Figure 2). The aixCAVE provides a five-sided IVE with a size of $5.25m \times 5.25m \times 3.30m$ (width × length × height). The five projection displays, i.e., the four walls and the floor, enclose the user giving her a 360 degree field of regard in the hori-

[9] The *uncanny valley* theory originates from the field of robot-human interaction (Mori, MacDorman, & Kageki, 2012), and has been transferred to virtual human-human research (Brenton, Gillies, Ballin, & Chatting, 2005). This theory states that as the quality of personification increases, a human's emotional response to a VH gets increasingly positive and emphatic, up to a point beyond which the response quickly turns into strong revulsion and then becomes positive again. Thus, the uncanny valley theory postulates that there is a non-trivial relationship between quality of personification and the acceptance of VHs.

zontal plane in the computer-generated environment (Kuhlen, 2014). In our setting, when a human subject enters the aixCAVE, she perceives the virtual environment as a production hall. Being inside this IVE, the human subject performs a sorting task at a virtual conveyor belt while seeing a virtual co-worker that independently performs the same task.

Figure 2: The aixCAVE at RWTH Aachen University



By employing a stereoscopic projection approach in conjunction with special 3D glasses, the virtual scenario experienced in the aixCAVE is presented in a three-dimensional, quasi-holographic way as known from 3D cinemas. Due to the high resolution, the user experiences a very detailed and crisp image of the virtual environment.[10] Images are generated through a user-centered projection, i.e., depending on the user's position and viewing direction. To this end, tracking markers are attached to the glasses, which inform the system about the user's head pose. With this, a user can see the scene from any desired perspective. The projector technology causes the aixCAVE to be a single-user system.

5.2 The task

We use a simple real effort task. A constant stream of objects (abstract cubes) transported on a conveyor belt approach the subject. To facilitate a natural way of interaction with the virtual environment, and an intuitive sorting of virtual cubes on a conveyor belt, the subject's hand has to be tracked. For this, a hand band with tracking markers on it is

[10] In total 24 projectors are used with a resolution of 1920×1200 pixels each, 4 per wall and 8 for the floor.

attached to the user's dominant hand.[11] Using the pose information of the user's hand, the human subject can pick up the virtual cubes and inspect them, and throw them into a bin, just by moving her hand.

The height of the conveyor belt could be easily adjusted, so each subject had a comfortable working position. There are two types of cubes. The first type of cubes has six blue-shaded sides, representing objects *without a defect*. The other type of cubes has five blue-shaded and one red-shaded side, they represent objects *with a defect*. The red side is not visible to the subject when the cube appears on the belt. The red side is located either on the bottom, the back or the right-hand side of the cube. The task is to sort out the objects with a defect (having a red side). To inspect a cube, the subject must grasp a cube and rotate it. If the cube is defect, the subject has to put it into a bin, or, if not defect, put the cube back onto the belt. In a series of pilot sessions, we adjusted the speed of the belt such that for an average subject it is challenging to inspect every cube in the given time.

We do not provide any financial incentives to work. There are no penalties for the defect cubes that were left unsorted on the belt. During the instructions, we told subjects that they should "reach a work output as good as possible".[12]

In the initial phase, 34 out of 168 cubes are defect and no co-worker is present. The the second phase (the peer phase) is twice as long (68 out of 336 cubes are defect) and a virtual co-worker is present and can be observed from behind. The order of cubes was randomly drawn before the start of the first session and is the same for all subjects and between treatments. Between the first and second phase, there is a short break, where the experimenter enters the VR lab and informs the subject about the second (peer) phase handing out a short note (see Appendix, subsection C.1). In particular, the subject learns that in the second phase there will be a "co-worker" who performs the same task as the subject but does so independently. The subject also learns that her co-worker is computer-controlled and represented by a VH.

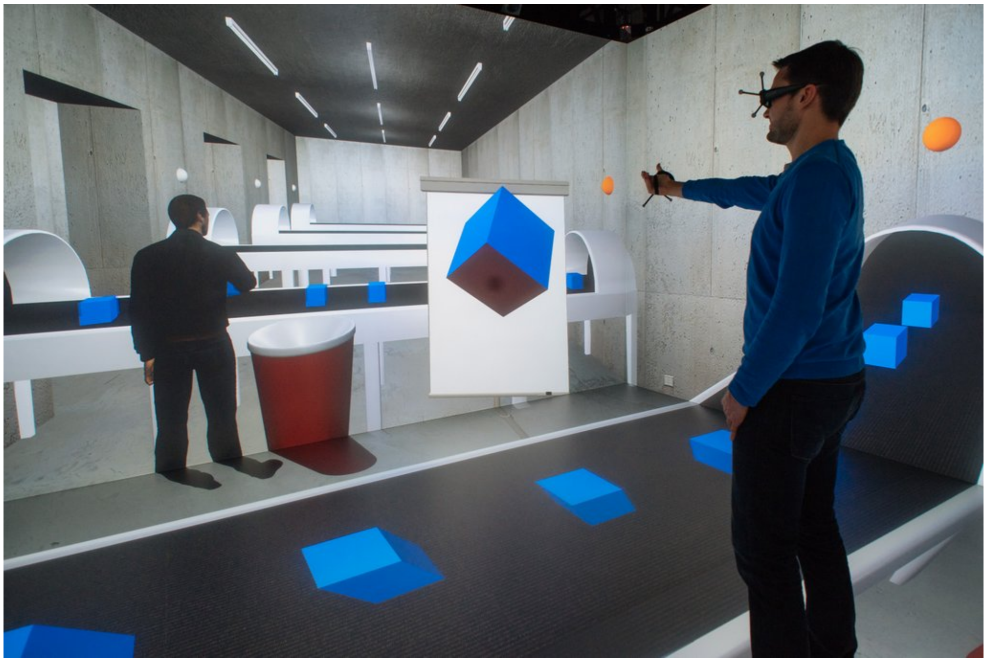
5.3 The virtual human as a peer

To make the VH move as naturally as possible, we captured a student assistant's movements via a Kinect 360 (Windows, 2011) while he performed the task in the aixCAVE. We transferred the data using 3ds max (Autodesk, 2014) into the model of a male VH, provided by SmartBody (Shapiro, 2011). Using this procedure, we modeled two otherwise identical virtual humans with different behaviors: one worker sorting virtually all cubes, and one

[11] We asked subjects for their handedness before starting the experiment and reversed the whole virtual environment for the left-handed. In particular, cubes approached left-handed subjects from the left, and right-handed subjects from the right.

[12] The original instructions were in German. An English translation is included in the Appendix C.

Figure 3: The experimental setup inside the aixCAVE



worker only sorting a subset of cubes while having short rests during the task.[13]

5.4 The experimental procedures

Subjects were recruited with ORSEE (Greiner, 2015). They were paid a fixed amount of 12 Euros. An experimental session including instructions and the questionnaire took approximately 45 minutes. The experimental software was developed by the Visual Computing Institute of RWTH Aachen University. Programming of the questionnaire was done using z-Tree (Fischbacher, 2007).

Table 1: Participant statistics

	Men		Women		Total	
Treatment	Obs	Mean Age	Obs	Mean Age	Obs	Mean Age
SLOW	37	25.2	17	23.4	54	24.6
FAST	29	26.2	25	22.6	54	24.6

The data was collected in 108 sessions, in each session one subject was acting in the aixCAVE. As can be seen in Table 1, we have 54 observations in each treatment. While 37 male subjects and 17 female participated in the SLOW treatment, we had 29 male and 25

[13]See the video clips on https://www.youtube.com/watch?v=RuX688VliB4 for the VH we modeled for the SLOW treatment, and https://www.youtube.com/watch?v=xZgBdlD1ypQ, for the FAST treatment, respectively.

female subjects in the FAST treatment.[14] The sessions were conducted on eight days. To avoid possible biased behavior emerging from different days and different times of the day, we constantly alternated the sessions between the SLOW and FAST treatment.

6 Results: Virtual humans and peer effects

We organize our results in two sections. In Section 6, we analyze performance data similar to those collected in related experimental studies. In Section 7, we perform additional analyses based on performance measures that use tracking data that is easily generated in IVEs.

6.1 Human subjects' perceptions

To evaluate the subjects' perception of the IVE including the VH and the task, we use a variant of the presence questionnaire developed by Slater, Usoh, and Steed (1994). Table 2 summarizes the results concerning subjects' general experience in the aixCAVE for the two different treatments. Subjects evaluated different aspects of the IVE on a 7-point Likert scale with high numbers indicating higher realism. In general, subjects evaluate their experience in the aixCAVE as being realistic, and for four of six categories the ratings in both treatments are very similar. Only in two categories, subjects in SLOW consider their "sense of being" and the realism of the IVE (weakly) significantly higher than those in FAST ($p = 0.078$ and $p = 0.012$, Mann-Whitney rank-sum test, two-sided)[15], see the first two questions of Table 2.[16]. The scores we observe are similar to those obtained in other CAVE studies (Usoh, Catena, Arman, & Slater, 2000).

For the experimenter, another way to get insights about subjects' perception of the IVE is to observe subjects' behavioral responses. How do subjects move in the IVE, and how do they react to the VH? Anecdotal observations of participants' movements in the aixCAVE suggest a realistic interaction with the virtual environment. Indeed, it happened that one participant of our experiment wanted to lean on the virtual conveyor belt and she fell on the ground.[17]

[14]We had to cancel another 11 observations due to software crashes.

[15]Unless otherwise stated, we use the non-parametric Mann-Whitney rank-sum test for between treatment comparisons, and report the p-values of two-sided tests.

[16]The numbers in Table 2 are similar to those in a related study conducted in the aixCAVE that did not involve VHs (DeHoratius et al., 2018). For example, "sense of being" was rated on average 5.53 and realism was rated 4.26. The remaining questions were rated with 4.28, 4.94, 4.01, and 4.42, respectively.

[17]Luckily, she did not hurt herself and could go on with the experiment.

Table 2: Subjects' experience in the aixCAVE

Treatment	Sense of Being	Virt. Env. Was Reality	Saw vs Visited	Where Were You	Memory Structure	Physically in Virt. Env.
SLOW	5.57	4.81	4.04	4.91	4.39	4.44
	(0.92)	(1.26)	(1.94)	(1.55)	(1.55)	(1.64)
FAST	5.13	4.07	4.31	4.87	4.50	4.30
	(1.32)	(1.58)	(1.85)	(1.61)	(1.75)	(1.89)

Note: All questions were measured on a 7-point Likert scale with higher numbers indicating a greater sense of realism in the virtual environment. Standard deviations in parentheses.

Table 3 summarizes subjects' perceptions of the VH (see Table 3). Participants were not able to exactly observe the virtual peer's productivity in terms of correctly or wrongly sorted cubes, simply because they were standing behind the VH with no clear vision on all sides of a cube in the virtual peer's hand. Nevertheless, subjects in the FAST treatment rated the VH's productivity as being considerably higher than subjects in the SLOW treatment (SLOW: 4.09 and FAST: 5.15, $p < 0.001$).

Table 3: Subjects' perceptions of the virtual human and the task

Treatment	VH is productive	VH moves naturally	Enjoyed CAVE	Enjoyed Task	Exerted Effort
SLOW	4.09	4.78	6.44	5.56	6.04
	(1.61)	(1.45)	(0.84)	(1.34)	(0.95)
FAST	5.15	4.76	6.39	5.41	5.78
	(1.50)	(1.33)	(0.88)	(1.45)	(1.24)

Note: All questions were measured on a 7-point Likert scale. Standard deviations in parentheses.

Result 1 *Subjects in the FAST treatment evaluate the VH's work performance significantly higher than subjects in the SLOW treatment.*

This shows that our intended treatment variation was experienced by subjects in the intended direction. To see whether the observed difference between the ratings can be attributed to possible differences in the perception of the VH's movements, we also asked subjects about how they considered the realism of the VH's movements. In both treatments, participants judge the VH's movements similarly (SLOW: 4.78 and FAST: 4.76). These numbers underline, that the perceived differences in the VHs' behavior cannot be attributed to the perceived differences in VHs' movements.

The questionnaire also reveals (see Table 3) that subjects highly enjoyed the experience of being in the aixCAVE (SLOW: 6.44 and FAST: 6.39, no significant differences between treatments). The task was also enjoyed, but less than the experience of the aixCAVE (SLOW: 5.56 and FAST: 5.41, again no differences between treatments). This indicates that performance differences between treatments cannot be attributed to differences in how exciting the CAVE experience was or to the degree of task enjoyment. The last item in Table 3 supports this congruence of perception, as it shows that subjects' self-judgment of their exerted effort level is rather high and not significantly different between treatments.

6.2 Peer effects

We measure a subject's *productivity* as the sum of the number of sorted cubes with a defect put into the bin, i.e., cubes that are correctly sorted, minus the number of cubes without a defect put into the bin, i.e., cubes that are incorrectly sorted.

Table 4 displays the averages of productivity in the initial phase and the peer phase, and the average relative productivity increase for both treatments. If we look at productivity in the peer phase, there is no significant difference between the SLOW and the FAST treatments (SLOW: 48.8, FAST: 46.1, $p = 0.155$). There is no difference in the productivity increase between both treatments, either ($p = 0.851$).

Table 4: Productivity

Treatment	Productivity in the initial phase	Productivity in the peer phase	% Relative productivity increase
SLOW	21.9	48.8	14.3
	(4.6)	(9.8)	(25.9)
FAST	19.6	46.1	22.0
	(5.0)	(9.8)	(42.1)

Note: Numbers in the table refer to treatment averages. Standard deviations are in parentheses. The number of cubes in the peer phase is twice as large as the number of cubes in the initial phase which has been considered in the relative increase in productivity numbers we report.

Recall that the initial phase was identical for all subjects, in particular, the VH was absent. Nevertheless, subjects in the SLOW treatment display significantly higher productivity in the initial phase than subjects in the FAST treatment (SLOW: 21.9, FAST: 19.6, $p = 0.032$). We do not have an explanation for this observation than it came by chance. Even though we have an uneven distribution of gender between both treatments, the difference cannot be attributed to gender. There are no significant differences between women's and men's average productivity, neither in the SLOW treatment (women: 21.8, and men: 21.9, $p = 0.911$), nor in the FAST treatment (women: 19.0, and men: 20.6, $p = 0.185$).

Regression analysis

To further investigate the determinants of productivity, we conducted a series of regressions. The dependent variable in all our models is productivity $(\text{Prod})_i$ in the peer phase. As independent variables, we include base productivity as observed in the initial phase $(\text{baseProd})_i$ to account for differences in base productivity, an indicator variable $(\text{FAST})_i$ for the treatment ($(\text{FAST})_i = 1$ if subject i is in the treatment with the high productive VH), and an interaction term between $(\text{baseProd})_i$ and $(\text{FAST})_i$ (model (I)). We also included several control variables of which none turned out to be significant (model (II)).

$$(\text{Prod})_i = \alpha + \beta \cdot (\text{baseProd}_i) + \gamma \cdot (\text{FAST}_i) + \delta \cdot (\text{baseProd}_i \cdot \text{FAST}_i) + \epsilon_i, \qquad (1)$$

As can be seen in Table 5 the coefficients of the variables in model (I) reveal that productivity in the peer phase positively depends on the productivity exerted in the initial phase. Furthermore, there is no significant treatment difference regarding the level of productivity, as both the coefficient for the treatment dummy FAST as well as the interaction term FAST $\times$ baseProd are not significant. This is reconfirmed in model (II). In further regressions we also included the controls from Table 2 and Table 3 without this affecting the findings. These findings are in line with Hypothesis 1, while they do not support Hypothesis 2.

Result 2 *Over all subjects, there is no significant difference in productivity between the SLOW and the FAST treatment.*

6.3 Do competitive subjects perform better?

As discussed in subsection 6.1, subjects acknowledge the presence of a virtual co-worker. To understand whether and how subjects react to the virtual peer, we asked subjects whether they cared about being more successful than the VH. 30 subjects (27.8 percent) were affirmative to this question. We refer to these participants who seem to be intrinsically motivated, as "competitive" subjects. Averaged over all subjects in both treatments, competitive subjects display significantly higher productivity than non-competitive subjects (50.8 and 46.1, $p = 0.023$), which supports Hypothesis 3a.

Hypothesis 3a is also supported by two alternative regressions with COMPETITIVE as a dummy variable, and with an interaction term. Models (III) and (IV) in Table 5 reveal that competitive subjects indeed display higher productivity, as the respective dummy variables show. Specifically, model (IV) reveals that for competitive subjects, the effect of base productivity on the productivity in the peer phase is (weakly significantly) less strong than for the non-competitive subjects. Model (IV) also shows that for men, the effect of being competitive is (weakly significantly) less strong on the productivity than for women.

Table 5: Determinants of the productivity (in the peer phase)

	(I)	(II)	(III)	(IV)	(V)	(VI)
baseProd	1.163	1.106	1.261	1.272	0.932	0.869
	(0.251)***	(0.261)***	(0.196)***	(0.197)***	(0.284)***	(0.291)***
FAST	2.614	-0.714			-13.305	-16.253
	(7.342)	(7.623)			(8.230)	(8.560)*
FAST $\times$ baseProd	-0.151	-0.068			0.592	0.725
	(0.342)	(0.353)			(0.387)	(0.401)*
COMP			16.053	16.429	-13.007	-12.301
			(7.406)**	(7.374)**	(12.758)	(12.851)
COMP $\times$ baseProd			-0.566	-0.568	0.670	0.661
			(0.342)	(0.341)*	(0.553)	(0.558)
FAST $\times$ COMP					46.338	46.283
					(16.004)***	(16.233)***
FAST $\times$ COMP $\times$ baseProd					-2.103	-2.104
					(0.747)***	(0.758)***
Age		0.063		0.109		0.177
		(0.148)		(0.142)		(0.142)
Male		-2.469		-2.949		-2.980
		(1.847)		(1.750)*		(1.767)*
Engineer		0.984		0.538		1.167
		(1.776)		(1.665)		(1.685)
Left-handed		5.896		5.340		4.493
		(4.331)		(4.171)		(4.121)
Constant	23.386	23.954	20.040	18.383	27.512	25.603
	(5.615)***	(6.638)***	(4.164)***	(5.523)***	(6.142)***	(6.771)***
Adjusted R^2	0.27	0.27	0.33	0.33	0.35	0.36
N	108	108	108	108	108	108

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

The next question concerns Hypothesis 3b, i.e., whether competitive subjects show a higher performance in the FAST treatment than in the SLOW treatment. If we look at the averages, we see that competitive subjects' productivity 49.2 in FAST while it is 45.3 in SLOW ($p = 0.123$, one-sided test because of the directed hypothesis). If we consider subjects' base productivities', for competitive subjects we see, that the relative average increase in productivity between the initial and the peer phase is 57.6 in FAST while it is 12.8 in SLOW ($p = 0.014$, one-sided). Models (V) and (VI) provide further support for Hypothesis 3b. The interaction term FAST × COMP shows that the difference in performance between competitive subjects and the other subjects in the FAST treatment is larger than in the SLOW treatment. The interaction term FAST × COMP × baseProd, however, shows that productivity (in the peer phase) of the competitive subjects in the FAST treatment is increasing less in their base productivity, as compared to the SLOW treatment.

Result 3

a) *Competitive subjects show significantly higher productivity than non-competitive subjects.*

b) *Competitive subjects show significantly higher productivity in the FAST treatment than in the SLOW treatment.*

7 Performance analysis using tracking data

One of the advantages of conducting experiments in IVEs is that one can collect tracking data. Such data can help to identify determinants of superior performance, and might measure the exerted effort more directly than a potentially noisy output measure. In our experiment, we collected data about the subject's hand position, and about the position and the orientation of the subject's head. In addition, we obtained data about the exact position of each cube when grabbed, and whether and how it was rotated. From the raw data, we can extract information about the position and movements of a subject, how often the subject grabs a cube, how long each grab (inspection) takes. We can further infer from the head position and orientation the gaze direction of the subject. Combined with the positional cube data, we can conclude, which sides of the cube were inspected by the subject.[18]

7.1 A measure based on hand & head tracking: careful inspections

Using the position and orientation of a subject's hand and head, we are able to detect (i) whether and how long a cube was inspected and (ii) which sides of a cube a subject

[18]There are some few studies that use tracking data in the context of economic decision-making. McCall and Singer (2015) show that interpersonal distance measures can predict financial behavior. In their experiment, first, subjects play several rounds of a trust game with two confederate players, one being "fair" and the other "unfair". After the trust game, entering a virtual room, human subjects can approach the VHs of the fair and the unfair player. After the virtual encounter, subjects can punish the other players, by investing a token for a deduction of three tokens on the punished player's account. The main results are: first, when being in the virtual room, subjects keep the fair player closer. Second, subjects who invest in the punishment of the unfair player are more likely to stand directly in front of those players.

actually inspected. We use this information to construct a novel performance measure: the number of carefully inspected cubes. This performance measure makes sense, if the quality of inspection is particularly important, e.g., because of overlooking defect parts/cubes in the supply chain can induce large follow-up costs, or if measures based on numbers of correctly sorted cubes are sensitive to random exogenous influences.

Definition 1 *An inspection is defined as a* ***careful inspection****, if (i) the inspected cube is a non-defect cube, i.e., a cube without a red side, and the subject inspected all three hidden sides and put the cube back on the belt, or (ii) if it is a defect cube, i.e., a cube with a red side, the subject saw the red side, and correctly placed the cube into the bin.*

In particular, an inspection conducted by the subject is *not* a careful inspection, if the subject puts the cube on the belt without inspecting (looking at) all of its hidden sides. Table 6 displays averages of several measures based on tracking for the peer phase: the number of total inspections, i.e., the number of total grabs a subject performed; the number of careful inspections as defined above; the inspection accuracy, i.e., the number of careful inspections divided by the total number of inspections.

Table 6: Inspections

	# Total inspections		# Careful inspections		% Inspection accuracy	
Treatment	Comp=NO	Comp=YES	Comp=NO	Comp=YES	Comp=NO	Comp=YES
SLOW	302.7	309.4	211.2	245.0	70.7	79.4
	(38.5)	(32.5)	(73.7)	(73.5)	(24.0)	(27.9)
FAST	300.7	304.0	202.9	232.0	68.7	76.1
	(31.9)	(32.5)	(80.5)	(56.3)	(22.5)	(14.8)
Total	301.6	307.4	206.6	240.2	69.5	78.2
	(34.8)	(32.0)	(77.1)	(67.0)	(26.1)	(19.8)

Note: Standard deviations in parentheses. Comp=NO (YES) refers to non-competitive (competitive) subjects.

The number of careful inspections correlates significantly ($p < 0.001$) and positively with productivity ($R^2 = 0.748$). We do not find any significant treatment differences, neither for the number of total inspections, nor for the number of careful inspections. Also the number of total inspections does not differ significantly between competitive and non-competitive subjects (aggregated over both treatments, $p = 0.423$).[19] However, competitive subjects perform more careful inspections (per person) than non-competitive subjects ($p = 0.034$), and their accuracy is also (weakly) significantly higher ($p = 0.089$). We also repeated

[19]Note that in Section 6.3, we have shown that the competitive subjects display a higher relative increase in productivity than the non-competitive subjects.

the regressions from Table 5 with the number of careful inspections (in the peer phase) as dependent variable and with careful inspections in the initial phase being one of the explanatory variables. While model (IIIa) in Table A.1 in the Appendix A does not provide support for Hypothesis 3a, model (Va) shows evidence supporting Hypothesis 3b.

Result 4 *With number of careful inspections as performance measure the evidence that competitive subjects show a higher performance (over both treatments) than non-competitive subjects is mixed. Competitive subjects have a (weakly) significantly higher inspection accuracy and display a higher performance in the FAST than in the SLOW treatment.*

Position data is recorded with timestamps, which enables us to determine the duration of movements. We observe that careful inspections take approx. 0.25 seconds longer than the other "uncareful" inspections (1.40 and 1.15 seconds, respectively), a Wilcoxon matched-pairs is highly significant ($p < 0.001$). Competitive subjects display (weakly) significantly shorter inspection times in careful inspections than non-competitive subjects (1.32 and 1.42 seconds, $p = 0.055$). Note that the combination of both position data and timestamps can also be used to determine the speed of the movements.

7.2 Body movements and productivity

This section demonstrates the potential of VR tracking data to learn more about the relationship of subjects' physical (body) movements and work performance. In our setting, subjects execute movements, like moving along the conveyor belt from one place to another, grabbing cubes, holding the cubes in their hands, inspecting the cubes by rotating. One can imagine these movements may affect productivity. For example, consider the position of subjects when grabbing the cubes. Should one chase every cube on the belt, or is it perhaps more effective to stick one position at the conveyor belt even if this implies missing out on some cubes? To answer this question, we use the position of the subjects' head in space when grabbing a cube.

Definition 2 *We define the **head distance** as the distance (in centimeters) between the point in space where the subject's **head** is located when the subject starts to handle a cube, and the point in space where her/his head was located when the subject started handling the previous cube.*

Another interesting body movement that could affect performance is the movement of the subject's hand between handling two consecutive cubes. One can imagine that a high productive worker moves his/her hand more efficiently/smoothly, when grabbing the next cube after having inspected the last one. For example, a high productive worker's hand distance may cover a shorter distance when grabbing the next cube than a low productive worker's hand.

Definition 3 *We define the **pausing distance** as the distance (in centimeters) the subject's hand covers between handling two consecutive cubes, i.e., the distance covered when the subject's hand releases a cube until the subject grabs the next cube.*

To see whether the two measures defined above are related to a subject's performance, we run several regressions. Models (I) and (II) in Table 7 refer to the head distance, while models (III) and (IV) refer to the pausing distance. While models (I) and (III) control for treatment differences, models (II) and (IV) investigate whether competitive subjects perform differently than non-competitive subjects, across treatments. The analysis from Section 7.1 suggests that competitive subjects may work differently than non-competitive subjects. This may be due to differences in the movements between both groups.

Model (I) reveals that head distance does not affect productivity significantly. Model (II) shows, that being a competitive subject increases productivity; in addition for the group of competitive subjects, there exists a (weakly) significant relationship between head distance and productivity: the less the movement of the head between two consecutive cubes, i.e., the smaller the head distance, the higher is the productivity of a competitive subject.

From model (III) we see that the pausing distance has a (weakly significant) negative impact on productivity: the more productive subjects display shorter pausing distances. Finally, model (IV) pools again the data over both treatments. For competitive subjects it shows, there is a significant negative relationship between pausing distance and productivity.

Result 5 *For competitive subjects, smaller moves of the body (measured as the head distance) and quicker handling of cubes (measured as shorter pausing distances) have weakly significantly positive effects on productivity.*

Table 7: Movements and productivity

	(I)	(II)	(III)	(IV)
baseProd	1.153	1.256	1.172	1.245
	(0.258)***	(0.196)***	(0.252)***	(0.195)***
Head Distance	-0.238	0.034		
	(0.156)	(0.152)		
Pausing Distance			-0.115	0.028
			(0.065)*	(0.063)
FAST	-0.378		-5.675	
	(7.654)		(9.133)	
FAST × baseProd	-0.151		-0.175	
	(0.354)		(0.349)	
FAST × Head Distance	0.260			
	(0.247)			
FAST × Pausing Distance			0.130	
			(0.094)	
Competitive		18.510		28.904
		(7.502)**		(8.941)***
Competitive × baseProd		-0.447		-0.487
		(0.344)		(0.337)
Competitive × Head Distance		-0.407		
		(0.228)*		
Competitive × Pausing Distance				-0.203
				(0.089)**
Constant	26.320	19.703	30.968	18.474
	(5.985)***	(4.308)***	(7.106)***	(5.290)***
Adjusted R^2	0.28	0.34	0.28	0.36
N	107	107	107	107

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

8 Conclusion and Discussion

We present a novel methodology to study peer effects in the workplace. By incorporating a virtual human (VH) in an immersive virtual environment (IVE) as a peer, we avoid the reflection problem: while the VH is immune to be influenced by the human subject, it may affect the human subject. We find that human subjects rate the IVE and the VH as natural.

We find that competitive subjects perform well in our setting, in particular in the FAST treatment. Whether and to what extent this is due to competitive pressure, though, remains unclear, since subjects were no provided with explicit feedback on productivity. Our experimental setting can easily be modified to study different possible sources of competitive pressure. To make the effect of competitive pressure more salient, one could add some displays showing the productivity of the virtual peer (Graff et al., 2018). Increasing the speed of the conveyor belt, or adding more/similar objects that have to be sorted could also increase competitive pressure. The findings of DeHoratius et al. (2018) indicate that increasing the speed of the conveyor belt, and increasing the similarity of objects can hamper productivity.

The applied methodology allows for interesting further investigations that go beyond this study. First, VHs can easily be adapted to display specific physical attributes, such as gender, age, height, or ethnicity, and therefore are suitable for investigating how a variety of different physical characteristics of a person may influence her peers. In addition, over the course of sessions, the VH can display the exact expression with a precision and consistency that human actors/confederates might not achieve. Second, tracking data could be used to address additional research questions based on subjects' movements, for example, one can measure interpersonal distances to evaluate (unconscious) behavioral patterns. Previous studies show that people keep greater distances to VHs with different ethnic looks (Dotsch & Wigboldus, 2008), or angry looking VHs (Bönsch et al., 2018).

Another potential application of the methodology is employee training in order to deal with stress, not only at a conveyor belt, or in tournaments but in many working environments. To give a real-world example, Walmart is already using simple (non-interactive) VR setups to prepare its employees for the most stressful sales day, the Black Friday (Robertson, 2017). Virtual experiments have already implications in fields where the experimental research started a decade ago. VR research has paved the way for some real-life applications in (vocational) training, (self-)therapy of some phobia, e.g., fear of heights.

In IVEs it can be simple to create counter-factual settings. One can conduct experiments that otherwise would be too expensive, or too dangerous, like exposing subjects to a physical risk; ethically not justifiable, like the famous trolley task-experiments; or simply impossible otherwise, like being embodied in another virtual person, or even in a virtual animal. Another interesting possibility of VR experiments is that one can see and "physically" interact with others in an IVE, the use of virtual agents enable retaining anonymity, which opens the door for a new class of experiments with bodily/physical interactions.

We believe that *virtual experiments* have great potential to enlarge the toolbox of the experimental economists, complementing classic lab and field experiments. Using the VR technology, researchers can create experiments with more natural contextual cues, but still having the advantages of a lab experiment, without losing any control. Tracking data such as body and hand movements can be used to conduct novel analyses which may help to discover unique insights.

References

Autodesk. (2014). 3ds Max. "http://www.autodesk.de/products/3ds-max/overview". ([Online; accessed 30-May-2015])

Azoulay, P., Graff Zivin, J. S., & Wang, J. (2010). Superstar extinction. Quarterly Journal of Economics, 125(2), 549–589.

Bailenson, J. N., Blascovich, J., Beall, A. C., & Loomis, J. M. (2003). Interpersonal distance in immersive virtual environments. Personality & Social Psychology Bulletin, 29(7), 819–833.

Bainbridge, W. S. (2007). The scientific research potential of virtual worlds. Science, 317(5837), 472–476.

Bandiera, O., Barankay, I., & Rasul, I. (2010). Social incentives in the workplace. Review of Economic Studies, 77(2), 417–458.

Biocca, F., & Levy, M. R. (1995). Virtual reality as a communication system. Communication in the Age of Virtual Reality, 15–31.

Blascovich, J., & Bailenson, J. (2011). Infinite reality: Avatars, eternal life, new worlds, and the dawn of the virtual revolution.

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). TARGET ARTICLE: Immersive virtual environment technology as a methodological tool for social psychology. Psychological Inquiry, 13(2), 103–124.

Bombari, D., Schmid Mast, M., Canadas, E., & Bachmann, M. (2015). Studying social interactions through immersive virtual environment technology: virtues, pitfalls, and future challenges. Frontiers in Psychology, 6.

Bönsch, A., Radke, S., Overath, H., Asche, L., Wendt, J., Habel, U., & Kuhlen, T. W. (2018). Social vr: Howpersonal space is affected by virtual agents' emotions. In Virtual reality (vr), 2018 ieee.

Bönsch, A., Wendt, J., Overath, H., Gürerk, Ö., Harbring, C., Grund, C., … Kuhlen, T. W. (2017). Peers at work: Economic real-effort experiments in the presence of virtual co-workers. In Virtual reality (vr), 2017 ieee (pp. 301–302).

Brenton, H., Gillies, M., Ballin, D., & Chatting, D. (2005). The uncanny valley: does it exist. In Proceedings of conference of human computer interaction, workshop on human animated character interaction. Citeseer.

Bursztyn, L., Ederer, F., Ferman, B., & Yuchtman, N. (2014). Understanding mechanisms underlying peer effects: Evidence from a field experiment on financial decisions. Econometrica, 82(4), 1273–1301.

Charness, G., & Kuhn, P. (2011). Lab labor: What can labor economists learn from the lab? In Handbook of labor economics (Vol. 4, pp. 229–330). Elsevier.

Cornelissen, T., Dustmann, C., & Schönberg, U. (2017). Peer effects in the workplace. American Economic Review, 107(2), 425–56.

Cruz-Neira, C., Sandin, D. J., & DeFanti, T. A. (1993). Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE. In SIGGRAPH '93: Proceedings of the 20th annual conference on Computer graphics and interactive techniques (pp. 135–142). New York, NY, USA: ACM.

Cummings, J. J., & Bailenson, J. N. (2015). How immersive is enough? a meta-analysis of the effect of immersive technology on user presence. Media Psychology, 0(0), 1–38.

DeHoratius, N., Gürerk, Ö., Honhon, D., & Hyndman, K. B. (2018). Execution failures in retail supply chains - a virtual reality experiment. Chicago Booth Research Paper No. 15-47.

Dotsch, R., & Wigboldus, D. H. (2008). Virtual prejudice. Journal of Experimental Social Psychology, 44(4), 1194–1198.

Falk, A., & Ichino, A. (2006). Clean evidence on peer effects. Journal of Labor Economics, 24(1), 39–57.

Fiore, S. M., Harrison, G. W., Hughes, C. E., & Rutström, E. E. (2009). Virtual experiments and environmental policy - frontiers of environmental and resource economics. Journal of Environmental Economics and Management, 57(1), 65–86.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. Experimental Economics, 10(2), 171–178.

Georganas, S., Tonin, M., & Vlassopoulos, M. (2015). Peer pressure and productivity: The role of observing and being observed. Journal of Economic Behavior & Organization, 117, 223–232.

Gerhards, L., & Gravert, C. (2017). Because of you i did not give up – how peers affect perseverance.

Graff, F., Grund, C., & Harbring, C. (2018). Competing on the holodeck: On the effect of virtual peers and heterogeneity in dynamic tournaments.

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with orsee. Journal of the Economic Science Association, 1, 114–125.

Gürerk, Ö., Bönsch, A., Braun, L., Grund, C., Harbring, C., Kittsteiner, T., & Staffeldt, A. (2014). Experimental economics in virtual reality. MPRA Paper No. 71409.

Gürerk, Ö., & Kasulke, A. (2017). Does virtual reality increase charitable giving? an experimental study. https://ssrn.com/abstract=3072002.

Harrison, G. W., Haruvy, E., & Rutström, E. E. (2011). Remarks on virtual world and virtual reality experiments. Southern Economic Journal, 78(1), 87–94.

Herbst, D., & Mas, A. (2015, 30 October). Peer effects on worker output in the laboratory generalize to the field. Science, 350(6260), 545–549.

Innocenti, A. (2017). Virtual reality experiments in economics. Journal of Behavioral and Experimental Economics, 69, 71 - 77.

Jackson, C. K., & Bruegmann, E. (2009). Teaching students and teaching each other: The importance of peer learning for teachers. American Economic Journal: Applied Economics, 1(4), 85–108.

Kandel, E., & Lazear, E. P. (1992, August). Peer pressure and partnerships. Journal of Political Economy, 100(4), 801–817.

Kasap, Z., & Magnenat-Thalmann, N. (2007). Intelligent virtual humans with autonomy and personality: State-of-the-art. Intelligent Decision Technologies, 1(1), 3–15.

Kaur, S., Kremer, M., & Mullainathan, S. (2010). Self-control and the development of work arrangements. American Economic Review, 100(2), 624–28.

Ku, G., Malhotra, D., & Murnighan, J. K. (2005). Towards a competitive arousal model of decision-making: A study of auction fever in live and internet auctions. Organizational Behavior and Human decision processes, 96(2), 89–103.

Kuhlen, T. W. (2014). Exploring Virtuality: Virtualität im Interdisziplinaren Diskurs. In S. Jeschke, L. Kobbelt, & A. Dröge (Eds.), (p. 133-147). Springer Fachmedien Wiesbaden.

Loomis, J. M., Blascovich, J. J., & Beall, A. C. (1999). Immersive virtual environment technology as a basic research tool in psychology. Behavior Research Methods, Instruments, & Computers, 31(4), 557–564.

Malhotra, D. (2010). The desire to win: The effects of competitive arousal on motivation and behavior. Organizational Behavior and Human Decision Processes, 111(2), 139–146.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. Review of Economic Studies, 60(3), 531–542.

Mas, A., & Moretti, E. (2009, February). Peers at work. American Economic Review, 99(1), 112–145.

McCall, C., & Singer, T. (2015, 12 February). Facing off with unfair others: introducing proxemic imaging as an implicit measure of approach and avoidance during social interaction. PloS one, 10(2), e0117532.

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. IEEE Robotics & Automation Magazine / IEEE Robotics & Automation Society,

19(2), 98–100.
Robertson, A. (2017). Walmart is training employees with a black friday vr simulator. "https://www.theverge.com/2017/6/1/15725732/walmart-strivr-vr-training-module". ([Online; accessed 20-Sep-2018])
Shapiro, A. (2011). Building a character animation system. In International conference on motion in games (pp. 98–109).
Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., ... Sanchez-Vives, M. V. (2006). A virtual reprise of the stanley milgram obedience experiments. PloS one, 1, e39.
Slater, M., Rovira, A., Southern, R., Swapp, D., Zhang, J. J., Campbell, C., & Levine, M. (2013). Bystander responses to a violent incident in an immersive virtual environment. PloS one, 8(1), e52766.
Slater, M., & Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality. Frontiers in Robotics and AI, 3, 74.
Slater, M., Usoh, M., & Steed, A. (1994). Depth of presence in virtual environments. Presence, 3(2), 130–144.
Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. Presence Teleoperators and Virtual Environments, 6(6), 603–616.
Thöni, C., & Gächter, S. (2015). Peer effects and social preferences in voluntary cooperation: A theoretical and experimental analysis. Journal Of Economic Psychology, 48, 72–88.
Trogdon, J. G., Nonnemaker, J., & Pais, J. (2008). Peer effects in adolescent overweight. Journal of Health Economics, 27(5), 1388–1399.
Usoh, M., Catena, E., Arman, S., & Slater, M. (2000). Using presence questionnaires in reality. Presence: Teleoperators and Virtual Environments, 9(5), 497–503.
Van Veldhuizen, R., Oosterbeek, H., & Sonnemans, J. (2018). Peers at work: Evidence from the lab. PloS one, 13(2), e0192038.
Windows, M. (2011). Kinect For Windows. "http://www.microsoft.com/en-us/kinectforwindows/". ([Online; accessed 30-May-2015])
Zimmerman, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. Review of Economics and Statistics, 85(1), 9–23.

A Additional regressions

Table A.1: Determinants of the careful inspections in the peer phase

	(Ia)	(IIIa)	(Va)
carefulinspection (init phase)	1.668	1.787	1.500
	(0.181)***	(0.157)***	(0.213)***
FAST	-15.602		-45.400
	(26.046)		(28.925)
FAST × carefulinspection (init phase)	0.211		0.541
	(0.257)		(0.288)*
COMP		42.000	-21.835
		(29.427)	(41.072)
COMP × carefulinspection (init phase)		-0.192	0.416
		(0.283)	(0.378)
FAST × COMP			128.972
			(59.467)**
FAST × COMP × carefulinspection (init phase)			-1.263
			(0.589)**
Constant	54.691	38.627	63.450
	(19.241)***	(14.413)***	(21.984)***
Adjusted R^2	0.65	0.66	0.67
N	106	106	106

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

We repeat the regressions from Table 5 with number of careful inspections as the performance measure. Models (Ia), (IIIa), and (Va) correspond to the models (I), (III), and (V) from Section 6.3, Table 5. In particular, we make use of the independent variable careful inspections in the initial phase as a determinant of the careful inspections in the peer phase. We observe that, analogous to the effect of base productivity on productivity in the peer phase, careful inspections in the initial phase affect highly significantly the number of careful inspections in the peer phase. Looking at model (IIIa), we see that competitive subjects' performance is not significantly different from non-competitive subjects' performance, which does not support our Hypothesis 3a. Looking at model (Va), we observe that competitive subjects in the FAST treatment display significantly higher performance than competitive subjects in the SLOW treatment, which supports our Hypothesis 3b. We also ran regressions including control variables, analogue to models (II), (IV), and (VI) in

Table 5, which remain all insignificant.

B The reflection problem and how we handle it

A major challenge in the empirical as well as the experimental research on peer effects is the reflection problem (Manski, 1993). Does agent i influence agent j, or vice versa? Similar to a person in a mirror chamber who cannot determine the causality of reflections she is seeing, simultaneous observation of peers in a group does not reveal who is influencing whom. This reflection problem severely hampers identification, if data is collected from real life or close to real life working environments, where, in principle, information is generated and received by subjects simultaneously. Identification is further complicated by the problem that (at least some) information between subjects flows freely and cannot be controlled. This is the case in field studies but also causes problems in lab experiments, such as in the study by Falk and Ichino (2006), where human subjects are physically located in the same room and can see each other, and each others' characteristics.

So far, the reflection problem has been addressed in lab experiments by controlling the information each subject receives on others' productivity. In Van Veldhuizen et al. (2018), a monitoring subject i is provided with productivity information on the monitored subject j, but not vice versa. Thöni and Gächter (2015) provide subjects with relevant information concerning their peers' performance only after they have provided effort themselves, and then – after having received the information on peers – allowing them to reconsider their own effort choices. Such limitations on information flow and decision making, however, add to the artificiality of the work situation and effort measurement that are typical for lab experiments.[20]

In this paper, we address these problems differently. We introduce virtual humans as subjects' peers. The virtual peer, though being a (fully controlled) computer program, resembles in its movements and appearance a human, and performs the same task as the human subject. Since the virtual peer is not influenced by the subject, we are able to observe non-confounded causal peer effects of the VH's working behavior on the human subject's performance. Obviously, the potential "cost" of this approach is that it is, per se, unclear whether subjects react similarly to humans and to VHs – though here, we build and rely on literature in social sciences and computer science that is mainly affirmative to this question (see Section 5.3).

[20]Virtually all lab experiments either use simple choices as a proxy for effort or tasks that are merely cognitive. In addition, in the lab, subjects only obtain numerical information about their peers' effort choices.

C Instructions (Translated from German)

Welcome to today's experiment in the aixCAVE. Please read the following instructions carefully. If you have a question, please ask the instructor. Please turn off your mobile phone!

Today's experiment consists of an **introduction**, a **testing part**, **main part**, and a concluding **questionnaire**. The introduction takes about 5 minutes and will familiarize you with how to execute today's **work task** in the aixCAVE. In the subsequent **testing part** you work exactly **5 minutes** without assistance. The introduction as well as the testing part are **not relevant for your payoff**. The **main part** starts immediately after the **introduction** and takes exactly **10 minutes**. You will work on the same tasks as in the testing part. Different from the testing part, you will receive a payment in the main part, as explained in detail below.

Your work task

Today, you will work at a conveyor belt. Your working area is the area in front of the conveyor belt. Your task is **to find as many defect objects as possible on the conveyor belt and to remove them**. In order to do so, you grasp the object and rotate it with your hand carefully. The objects are represented as blue cubes. The object is **OK (non-defect)** if all **six sides** of the cube are **blue**. If **one** side of the cube is colored **red**, than the object is **defect**. **You have to take away a defect object from the conveyer belt and to put it into the bin that is designated for this object**. Non-defect objects must remain on the conveyor belt. The **work output** measures your performance in this task.

Your work output

The following apply:

- Each **defect** object you put into the bin **increases** your work output by 1 unit.
- Each **non-defect** object in the bin **decreases** your work output by 1 unit.

Introduction

- You work several minutes on the task described above.
- You should get used to the (technical) implementation of the task.
- You can ask the present instructor questions.
- You receive for this part of the experiment no payoff.

Testing Part

- You work exactly 5 minutes on the task described above.
- You should achieve a working result as high as possible.
- You receive for this part of the experiment no payoff.

Main Part of the experiment Relevant for payoff

- You work exactly 10 minutes on the task described above.
- You should achieve a working result as high as possible.
- You receive for this part of the experiment a payoff.
- You will be informed about your payoff at the beginning of the main part.

The Questionnaire

- A short questionnaire follows.
- You will receive your payoff from the experiment in cash, after completing the questionnaire.

Please follow the points below while you are inside the aixCAVE:

- Please enter the aixCAVE with the slippers provided here.
- Please avoid quick or hectic movements in the aixCAVE.
- Please do not touch the walls or the floor of the aixCAVE.

We wish you success!

C.1 Information given to the subjects before the Treatment phase starts

The main part of the experiment begins!

- You work exactly 10 minutes on the task.
- You should achieve a work output as high as possible.
- You will be payed for this part of the experiment.
- Your payoff is 12 Euro, independent of the level of your work output.

In the main part, you have a co-worker who works at the same task as you do!

- Your co-worker is computer programmed.
- You both work independent of each other. This means in particular, that your work output does not depend on your co-worker's work output, and vice versa.

D Questionnaires

D.1 Questions about subjects' experience in the CAVE

The following contains detailed information on the questions reported in Table 2 about subjects' experience in the virtual environment. In all cases, the measure is a 7-point Likert scale with higher numbers indicating a greater sense of realism in the virtual environment. The questions were adapted from slater1994depth.

1. **Sense of Being:** Please rate your sense of being in the virtual environment, on a scale of 1 to 7, where 7 represents your normal experience of being in a place. I had a sense of "being there" in the virtual environment *not at all* (1) – *very much* (7).

2. **Virtual Environment Was Reality:** To what extent were there times during the experience when the virtual environment was the reality for you? There were times during the experience when the virtual environment was the reality for me *at no time* (1) – *almost all the time* (7).

3. **Saw vs. Visited:** When you think back to the experience, do you think of the virtual environment more as images that you saw or more as somewhere that you visited? The virtual environment seems to me to be more like *images that I saw* (1) – *somewhere that I visited* (7).

4. **Where Were You/Sense of Location:** During the time of the experience, which was the strongest on the whole, your sense of being in the virtual environment or of being elsewhere? I had a stronger sense of *being elsewhere* (1) – *somewhere that I visited* (7).

5. **Memory Structure:** Consider your memory of being in the virtual environment. How similar in terms of the structure of the memory is this to the structure of the memory of other places you have been today? By "structure of the memory" consider things like the extent to which you have a visual memory of the virtual environment, whether that memory is in color, the extent to which the memory seems vivid or realistic, its size, location in your imagination, and other such structural elements. I

think of the virtual environment as a place in a way similar to other places that I've been today *not at all* (1) – *very much so* (7).

6. **Physically in Virtual Environment:** During the time of your experience, did you often think to yourself that you were actually in the virtual environment? During the experience, I often thought that I was really standing in the virtual environment *not very often* (1) – *very much so* (7).

The following contains detailed information about the questions reported in Table 3 about subjects' perceptions with the treatments. In all cases, the measure is a 7-point Likert scale with higher numbers indicating greater agreement with the question.

1. **Enjoyed Cave:** I enjoyed the experience in the cave. ((1) Do not agree at all; (7) agree fully)

2. **Enjoyed Task:** I enjoyed the execution of the task. ((1) Do not agree at all; (7) agree fully)

3. **Exerted Effort:** I exerted much effort in the execution of the task. ((1) Do not agree at all; (7) agree fully)

D.2 Questions about subjects' perception of the virtual human

The following contains detailed information about the questions reported in Table 3 about subjects' perceptions of the VH.

1. **Virtual Human's productivity:** What do you think about the VH's work output. ((1) It was not good at all; (7) it was pretty good)

2. **VH moves naturally:** Please give a rating, to which extent the VH moves naturally. ((1) Not naturally at all; (7) pretty much naturally)

D.3 Questions about subjects' willingness to compete with the virtual human

1. **Success:** Was it important to you to be more successful than the VH. ((0) No; (1) Yes)