Master Thesis

31 January, 2021

Fostering Productivity and Wellbeing in Team Settings

A holistic approach for supporting software developers.

Anastasia Ruvimova

of Halle, Germany (15-720-063)

supervised by

Prof. Dr. Thomas Fritz

University of Zurich UZH

s.e.a.l. software evolution & architecture lab

Master Thesis

Fostering Productivity and Wellbeing in Team Settings

A holistic approach for supporting software developers.

Anastasia Ruvimova

University of Zurich[UZH]

s.e.a.l. software evolution & architecture lab

Master Thesis

Author: Anastasia Ruvimova, anastasia.ruvimova@uzh.ch

Project period: August 15, 2020 - February 4, 2021

Software Evolution & Architecture Lab
Department of Informatics, University of Zurich

Acknowledgements

I would like to take this opportunity to thank everyone involved in my thesis and in my journey towards a Masters: my advisor, Prof. Dr. Thomas Fritz for his constant support, Prof. Dr. Elaine Huang and members of the HASE group for the ready assistance and enlightening discussions, my parents for their steadfast belief in me, and all of the friends and colleagues who have gently pushed me along the way. Without you, I would not be where I am today.

I would further like to thank all study participants and assistants involved in our studies. We thank the Natural Sciences and Engineering Research Council of Canada (NSERC), the Saskatchewan-Waterloo Games User Research (SWaGUR) program, and the Ontario Ministry of Research, Innovation and Science for funding. We would also like to thank Willamette University for providing a space to run our study.

Abstract

The modern workplace demands more of its employees than ever before as technology assumes a larger role. In this context, the productivity and wellbeing of the individual as well as the team and the overall organizations is increasingly important, yet more research is needed to understand them and their relation. In a set of two research studies, we explored the productivity and wellbeing of software developers in team dynamics. We first examined the effects of virtual and traditional work environments on knowledge workers while performing work-related tasks. The qualitative and exploratory quantitative results show that the closed office and the beach VR are similarly good according to users' ranking in reducing distractions and inducing flow, and that these two environments are preferred over the non-VR open office and VR open office environments. Overall, these results indicate the potential that VR environments have to help knowledge workers achieve flow and stay calm and focused even in loud open office work settings. Second, we designed and piloted a multi-month field study in which participants regularly respond to a series of productivity and wellbeing questionnaires throughout their workday. The multi-tiered data collection includes hourly, daily, weekly, monthly, and one-time surveys for a better understanding of the fine- and coarse-grained factors which affect the productivity and wellbeing of software teams. The broader vision of our research is to better understand productivity and wellbeing on an individual and team level, and to develop approaches that support professionals in spending their time well at work. By taking a more holistic approach and incorporating developer wellbeing, we can design more effective strategies for an upwards cycle of productivity and wellbeing.

Disclaimer

The first study presented in this thesis (chapter 3) was completed in collaboration with Prof. Dr. Thomas Fritz (University of Zurich), Prof. Dr. Mark Hancock (University of Waterloo), Prof. Dr. David Shepherd (Virginia Commonwealth University) and Junhyeok Kim (University of Waterloo). Our paper was published in the proceedings of the ACM SIGCHI Conference in Honolulu, Hawaii, April 2020 [94].

The second study (chapter 4) was designed in collaboration with Prof. Dr. Elaine Huang, Prof. Dr. Lauren Howe, Claudia Vogel, Jan Gugler, and Alexander Lill.

Contents

Chapter 1

Introduction

This fable is the story of a poor farmer who one day discovers in the nest of his pet goose a glittering golden egg. At first, he thinks it must be some kind of trick. But as he starts to throw the egg aside, he has second thoughts and takes it in to be appraised instead. The egg is pure gold! The farmer can't believe his good fortune. He becomes even more incredulous the following day when the experience is repeated. Day after day, he awakens to rush to the nest and find another golden egg. He becomes fabulously wealthy; it all seems too good to be true.

But with his increasing wealth comes greed and impatience. Unable to wait day after day for the golden eggs, the farmer decides he will kill the goose and get them all at once. But when he opens the goose, he finds it empty. There are no golden eggs — and now there is no way to get any more. The farmer has destroyed the goose that produced them.

But as the story shows, true effectiveness is a function of two things: what is produced (the golden eggs) and the producing asset or capacity to produce (the goose).

If you adopt a pattern of life that focuses on golden eggs and neglects the goose, you will soon be without the asset that produces golden eggs. On the other hand, if you only take care of the goose with no aim toward the golden eggs, you soon won't have the wherewithal to feed yourself or the goose.

Effectiveness lies in the balance — what I call the P/PC Balance. P stands for production of desired results, the golden eggs. PC stands for production capability, the ability or asset that produces the golden eggs.

- Stephen Covey

1.1 In Search of Productivity

In today's fast-paced world, the strive for productivity can resemble a wild race for golden eggs. This buzzword attracts many knowledge work spheres: companies clamor to lay their hands on the latest tools and tricks which will enable their employees to become faster, more efficient, and more profitable. A wealth of research is dedicated to exploring and mining productivity (see Chapter 2). In this research, productivity is often defined in terms of output per time. And yet, is this the right approach? And is it even profitable in the long run? Research has shown that while in the short-term, high workload can lead to a sense of accomplishment and boosted morale, sustained overworking can result in burnout and fatigue.

While employers may profit in the first few months of increased output, they may find that in the end, the developers grow fatigued and their productivity drops even below their regular levels. This concept is illustrated by Aesop's fable of the Golden Goose, as retold by Stephen Covey. While companies may succeed in a short-term boost of work output (golden eggs), they will eventually wear out the productivity capacity of the worker (the golden goose).

Recently more related work has therefore focused on the developers' perception of productivity [81, 82] and the factors affecting productivity, including developers' feelings [46, 84, 110], yet mostly focusing on the individual and neglecting the team context in which developers work. In our research, we extend previous work and choose to take a more holistic approach by incorporating wellbeing, team context, and other soft factors in the understanding of productivity and what effects it. Our research objective is to foster productive work, taking into account individual and team dynamics.

In this thesis, we present two recent studies which approach this objective from slightly different angles. As a first step, we look at the potential of novel technologies to boost productivity in a team context. Specifically, we examine the use of Virtual Reality (VR) technology to foster flow in open office team environments. Our hypothesis is that using VR to simulate a calming environment can reduce interruptions, and improve flow as well as wellbeing. At the same time, we are interested in a more broad examination of the factors affecting productivity and wellbeing from an individual and team perspective. Though individual productivity and team productivity have been investigated separately, we have yet to understand the correlation between them. Therefore, as a second step in our research, we examine productivity of software development teams more holistically, including the different levels of productivity—-team and individual–their interplay and their relation to wellbeing in the workplace.

1.2 Contributions

This thesis contributes to the growing body of research on individual and team productivity. We present two studies where we examine different aspects. The first, a controlled lab study with 35 participants, investigates the effects of virtual reality on a person's flow and affective states when performing programming tasks. The second study, which has been designed and piloted, examines the relationship between individual and team productivity and wellbeing. The larger aim of our research is to harness the power of technology to boost the effectiveness of knowledge workers through smart practices and innovative tools.

1.2.1 Study 1: Boosting Flow through Virtual Reality

In a lab study, 35 participants performed visual programming tasks in four combinations of physical (open or closed office) and virtual environments (beach or virtual office). While participants both preferred and were in flow more in a closed office without VR, in an open office, the VR environments outperformed the no VR condition in all measures of flow, performance, and preference. The qualitative and exploratory quantitative results show that the closed office and the beach VR are similarly good according to users' ranking in reducing distractions and inducing flow, and that these two environments are preferred over the non-VR open office and VR open office environments. Further studies need to be run to examine the generalizability of our quantitative results.

Overall, these results indicate the potential that VR environments have to help knowledge workers achieve flow and stay calm and focused even in loud open office work settings. At the same time, the results open up new opportunities for research. Since the preferred work environment is highly individual, we need to examine which environment is best suited for which situation. Considering the high customizability of VR and the nearly infinite possibilities for environments, we may be able to tailor the VR work setting to individual preferences, the task, and possibly even the current mental state of the knowledge worker to provide the best experience in every moment. At the same time, we need to explore how this technology will alter workplace dynamics and social interactions among knowledge workers of the future, due to the current limitations of the technology.

1.2.2 Study 2: Productivity and Wellbeing in Teams

The modern workplace demands more of its employees than ever before as technology assumes a larger role. Workers have to perform a broad variety of complex tasks, working in a global and fast-paced environment, and experiencing interruptions and distractions. In this context, the productivity and well-being of the individual as well as the team and the overall organizations is increasingly important, yet little is known about them and their relation. We designed and piloted a 6 to 8 week field stud using a mixed-methods approach to capture both finer-grained and longer-term productivity patterns, collecting data on productivity–team and individual level–and on wellbeing. Specifically, we designed and developed approaches to collect fine-grained experience samples on productivity and wellbeing, daily and weekly diaries of work and productivity patterns, interviews to gather more in-depth qualitative data from participants on the study period, and a survey to collect demographic and personality data. Based on the successful pilot, we have now extended this study to a larger set of participants. We are currently running this study with more than 70 knowledge workers of 15 teams. The broader vision of our research is to better understand productivity and well-being on an individual and team level, and to develop approaches that support professionals in spending their time well at work.

Our thesis is structured as follows: we begin with an overview of related work in the area of productivity and wellbeing in chapter 1, then present our two studies in chapters 2 and 3, we then briefly discuss this research and directions of future work in chapter 4, and conclude with a summary in chapter 5.

Chapter 2

Related work

Researchers have examined developer productivity from many angles, technical and subjective, individual and team, and investigated the factors which affect it. Since interruptions are considered one of the major impediments to productivity, especially in collaborative work environments, there is also a large body of research that examines interruptions and the relation to productivity. Below we present the highlights of existing literature as it is relevant to the different aspects of our work, including research on flow and the use of Virtual Reality for work.

Quantifying productivity

Companies in recent years have been increasingly investing in the productivity of their employees. Programs, trainings, workshops - it's all about productivity. Yet what is productivity and how is it achieved? In its most basic terms, productivity is commonly defined as the ratio of outputs over inputs, where outputs may be, for example, product volume and input may be cost or manhours [88]. This definition may have sufficed for factory personnel in the Industrial Age, but for today's Information Era these calculations become more complex. When it comes to uncertainty, creativity, autonomy, and abstractness, knowledge work highly differs from the nature of traditional factory labor [3, 32, 37, 38]. For a software developer, these outputs and inputs become rather murky. In researching developer productivity, several approaches have been employed in an effort to quantify development work. These include number of lines of source code written (SLOC) in a set time frame [34, 111], the number of tasks [116] or function points [2, 63] completed per month, or change requests fulfilled [25, 83]. See [81] for a more complete list of quantifying productivity with technical factors.

Perceived Productivity

These seemingly objective measures, however, only truly capture a small part of a developer's work [107]. This is one factor which motivates us to seek a more holistic approach in defining productivity. Others argue against measuring productivity at all [68], especially so because people tend to optimize for whatever metric is being used – something known as Goodhart's law [26, 43]. With technical approaches, productive work can often be underestimated: for example, a worker developing new skills or mentoring a new hire does not always bring about immediate tangible output, and time spent in such pursuits can be significantly undervalued. Much of the community therefore chooses to focus on the perceived productivity of developers, and what affects it [66, 81, 82]. A recent study by Beller et al. attempts to bridge the gap between these two schools of thought (automated measure and perceived) with an empirical study measuring both, while understanding the factors which lead to this variance [17]. Past research has shed light on a vast array of factors potentially affecting productivity. Personal factors include such things as intrinsic motivation [52], psychological well-being [27, 105, 113], and work engagement [12, 114], while organizational factors include job characteristics [22], feedback [36], autonomy [36, 101], and office environments [27] as affecting productivity. For a further review, see [54, 84, 110].

Predictive Factors of Productivity

Notably, much research identifies soft factors as playing a role in perceived productivity. For example, aspects of the job which motivate developers or bring enjoyment were shown to lead to higher productivity and job satisfaction [16, 66]. Perceived satisfaction levels of knowledge workers [77], as well as perceived affective states of software developers [46] were further shown to be correlated with productivity and efficiency. Likewise, developer's moods were shown to influence performance on certain tasks, such as debugging [64]. Self-reported satisfaction levels of knowledge workers [77], and more specifically, self-reported affective states of software developers [46], have further been shown to be strongly correlated with productivity and efficiency. Similarly, developers' moods have been shown to influence developers' performance on performing programming tasks, such as debugging [64]. We hope to add to this body of research by better understanding the relationship between productivity and wellbeing, and how to foster and balance both in an effective workflow.

Team Productivity

The research community has examined not only factors affecting an individual's productivity, but the team's as well. Software development is a collaborative endeavour and several aspects have been shown to potentially predict the success of software teams. In a survey of 31 software teams, Lakhanpa et. al found team cohesion to be the dominating factor when investigating the influence of team cohesion, team experience, and team capability on team performance [70]. A recent study of teams working remotely (due to the Pandemic) found that some predictive factors of productivity included the ability to brainstorm with colleagues, having less awareness of what colleagues are working on, and having difficulty communicating with colleagues, and these factors all have a significant relationship with changes in team productivity. Other factors include satisfaction with social interaction from social activities, while communication breakdowns on teams are also important factors when modeling change in team productivity [90]. While individual and team factors have been examined separately, there is a gap in understanding how an individual's perceived productivity affects the team's and vice versa. In Study 2, we design a study to investigate this relationship.

Productivity in the Pandemic

The COVID-19 pandemic has dramatically changed the day-to-day of many software teams. Those accustomed to collocated working, especially, have had to adapt to distributed and often asynchronous working formats. Ralph et al. [90] summarizes research on work in the pandemic. Early evidence suggests complicated effects on productivity, which vary by person, project and metric [13]. Some research indicates that programmers are working longer hours, at an unsustainable pace [42]. Working from home, by itself, is a major shift for many used to the office space. Generally, working from home is often claimed to improve productivity [24, 33, 79] and teleworkers consistently report increased perceived productivity [15, 39]. Interestingly, Baker et al. found that organizational and job-related factors (e.g. management culture, human resources support, structure of feedback) are more likely to affect teleworking employees' satisfaction and perceived productivity than work styles (e.g. planning vs. improvising) and household characteristics (e.g. number of children) [11] . While increasing productivity, "working from home is associated with greater levels of both work pressure and work-life conflict" [93, p. 92] because work intrudes into developers' home lives through working unpaid overtime, thinking about work in off hours, exhaustion and sleeplessness [57]. Moreover, individuals' wellbeing while working remotely is influenced by their emotional stability (that is, a person's ability to control their emotions when stressed). For people with high emotional stability, working from home provides more autonomy and fosters wellbeing; however, for employees with low emotional stability, it can exacerbate physical, social and psychological strain [87]. The COVID-19 pandemic has not been good for emotional stability [8]. Research on working from home has been criticized for relying on self-reports of perceived productivity, which may inflate its benefits [10]; however, objective measures often lack construct validity [91] and perceived productivity correlates well with managers' appraisals [14]. In the ongoing Study 2, our participants are almost exclusively working from home, and many are for the first time dealing with the struggles which come with it. Many impediments to both productivity and wellbeing are exacerbated in this pandemic time. This makes our research all the more relevant: Which factors are predictive of developer productivity and wellbeing, and how do these perceptions differ on an individual level versus a team level?

Interruptions

Whether at home or in the office, interruptions are a perpetual nuisance in collaborative environments such as the software team. Developers work in teams and often rely on communication and exchange of knowledge to move forward in their work. Yet, unless well synchronized, this can often lead to being interrupted in the middle of a task. This is especially true for open offices, which have become increasingly popular due to their cost-effectiveness. A wealth of research [4, 5, 9, 58–60, 71, 74, 75, 100, 117] has reached the general consensus that interruptions hinder productivity. Specifically, people can take over 25 minutes to get back to work after a brief interruption [74], people tend to make more errors in work after even a brief interruption [5, 71] and take more time to make decisions [100]. Some studies have also explored the behaviour of people when interrupted and observed that they sometimes delay responding to distractions in order to set up the state of the task they intend to resume later [60], but that interruptions increase task completion time, as it takes time to resume the original task [4, 59, 60], and this increase is particularly detrimental for knowledge work [58]. While multitasking, interruptions have been found to occur on average 7 times per hour, mostly from email and Instant Messaging alerts (circa 2007) [60]. A related study found that more task switching led to a higher chance of getting interrupted [74]. In some cases, interrupted work is completed in the same time as non-interrupted work, but with higher stress (mental cost) [75] due to annoyance, frustration, and anxiety interruption [9, 117]. While there has been significant work on understanding how problematic interruption is and its impact on productivity, our first presented study adds to this body of literature in exploring the use of Virtual Reality (VR) to address interruptions in an open office environment and foster productive work. We ask the question: Can VR take away distractions and result in a similar flow experience to that of a closed office? See chapter 3, where we explore this question in a controlled lab study.

Flow

In reducing distractions, we can help developers foster flow: a psychological state of full engagement and immersion in a task [30]. This affective state of complete absorption has been related to notions of peak performance and peak experience [89]. Several studies found that certain conditions are necessary to reach a state of flow, including balance between challenge and skill [28], the opportunity to learn [30], and self-control within the task [31]. Confidence and concentration have also been shown to facilitate flow [104]. Flow is also associated with higher performance [40] as well as higher quality of performance [104]. Flow, as we see, has the potential to facilitate both productivity and wellbeing.

Virtual Reality and Work

Researchers have experimented with a number of technologies with the intent of reducing distractions and fostering flow. One such novel technology with the potential to induce flow is Virtual Reality (VR). There are many practical applications of virtual reality, such as 3D modeling, education, and medicine [23, 53, 73, 98, 103]. While these could already be considered "work" applications, they are typically specialized uses of the technology to improve some aspect of the work practice, rather than a replacement for more generic desktop work. There have, however, been some attempts to bring 2D windows into 3D virtual environments (e.g., [6, 19, 41]), including commercial applications (e.g., Steam's Virtual Desktop). Other work has explored the use of keyboards for common desk work in offices [48, 67, 96], using VR for mobile knowledge work [49], and the effects of long-term use of VR in office environments [51].

While the possibility of office work in VR has been explored from a technical perspective, there has been little work exploring the use of this kind of interaction in VR to mitigate distraction in offices. One notable exception is the work by McGill et al. [78] which provides evidence that increasing awareness of the real world increases distraction from the VR experience, and suggestions for how much reality to include in VR. Some work has also shown that VR has the potential to help reduce stress, for example by simulating nature [108] or by facilitating meditation [106]. In Study 1, we build on this prior work by specifically targeting office work (programming) and leveraging VR to mitigate distractions and improve flow.

Through our research, we hope to understand productivity more holistically and to develop approaches and strategies which take into account individual and team dynamics in fostering an effective workflow for software development.

Chapter 3

"Transport Me Away": Fostering Flow in Open Offices through Virtual Reality

Open offices are cost-effective and continue to be popular. However, research shows that these environments, brimming with distractions and sensory overload, frequently hamper productivity. Our research investigates the use of virtual reality (VR) to mitigate distractions in an open office setting and improve one's ability to be in flow. In a lab study, 35 participants performed visual programming tasks in four combinations of physical (open or closed office) and virtual environments (beach or virtual office). While participants both preferred and were in flow more in a closed office without VR, in an open office, the VR environments outperformed the no VR condition in all measures of flow, performance, and preference. Especially considering the recent rapid advancements in VR, our findings illustrate the potential VR has to improve flow and satisfaction in open offices.

3.1 Introduction

For knowledge workers, especially software developers, flow [31]—characterized by energized focus and complete engagement on a given task—usually leads to increased productivity and personal development, even improving life satisfaction. While achieving flow is desirable, it can be challenging, as it requires an environment, that is "free from distractions" [95] for long periods of time [44, 85]. Unfortunately, this particular condition is becoming harder and harder to realize, as modern office life is fraught with distractions. Text messages, colleagues' questions, meetings, nearby phone conversations, and myriad other stimuli bombard office workers with constant interruptions [5, 74].

Because distractions are known to reduce productivity, it seems natural that employers would provide their knowledge workers with closed offices to minimize distractions. Why then are many employers moving to the open office concept, especially when it is known to cause distractions [21]? Unfortunately, because open offices are much cheaper than traditional, private offices, there is a clear incentive for employers to move to open offices. "Managers present [moving to open offices] as necessary for greater collaboration and productivity, but 99% of the changes are really driven by the desire to cut costs" [109].

Fortunately, at the same time that knowledge workers are being forced into distraction-filled open offices, virtual reality hardware is rapidly developing, presenting a potential solution. What if employers could continue to implement open office layouts, which save money, while, at the same time, provide a *virtually* closed office that offers employees many of the benefits of a traditional closed office?

Until recently, this concept was a cyberpunk [56] fantasy; headset resolution was low, causing text to be difficult to read [47], and headset tracking was poor, causing motion sickness in many participants [65]. However, recent advances have not only addressed these shortcomings, they have been dramatic enough to show that, within a few years, VR technology may improve drastically. For instance, the recently released HP Reverb offers the unprecedented resolution of 2160×2160 pixels per eye, making it possible to comfortably read small text within a headset. Furthermore, a recent review of studies in cybersickness (i.e., VR-induced motion sickness) show that adjustments to field of view along with a stationary setting (as would be used in a virtual office) can dramatically reduce motion sickness [92]. While the authors are optimistic that motion sickness can be addressed, work in this area of research is ongoing, and we would like to acknowledge that our suggested solution is dependent on solutions to known issues, such as sex differences [102] and further study of the phenomenon is required, though not the topic of this paper.

While virtual reality hardware is not yet ready for knowledge workers to don headsets full-time, it may be soon, and this work aims to investigate a future where it is viable. To investigate the potential feasibility of providing the advantages of closed offices via a VR-based office, we conducted a user study to compare several working modalities. We asked participants to complete tasks designed to simulate knowledge work in four different settings: a traditional closed office, a traditional open office (with distractions), a VR-based closed office (with distractions in the real world), and a VR-based open office (with distractions in the VR world). By comparing participants' performance and collecting their preferences, we have gained insight into the trade-offs that these environments offer. While participants performed more tasks more quickly in the traditional closed office than in either open office setting, there was only a small difference between performance in the traditional and VR closed offices. Similarly, while participants had the strongest preference for the traditional closed office, the VR closed office had a similar usability score, and both closed offices were strongly preferred over both open offices. We believe these results show potential that, as VR hardware advances, some of the benefits of closed offices could be brought to open office configurations via VR.

In our work, we leverage existing literature connecting flow and performance by comparing people's ability to get into a state of flow in open office settings and whether the use of virtual reality can help people to achieve this flow state. We hypothesize that higher flow can be achieved in open office settings when VR is used to mitigate distraction.

3.2 Study

We conducted a laboratory study to investigate the ability for virtual reality to help deal with distraction in open office environments. Specifically, we were interested in the following research questions:

RQ1 Does the use of a VR headset to perform work in an open office environment improve flow?

RQ2 Does the virtual scene rendered in the VR world matter, or will any environment do (even one with an open office simulation)?

RQ3 How close to closed office work can work in VR get?

Our study therefore included four environments to complete work: a traditional open office (baseline), a VR world with a tropical beach background, a VR world with an open office background, and a traditional closed office (the ideal being strived for).

3.2.1 Participants

We recruited 11 participants in the USA and 24 in Switzerland. We advertised using posters, university mailing lists, and university student job postings. The total pool of 35 participants is nearly balanced across two genders, with 18 identifying as women and 17 identifying as men. Because sex differences in spatial

Figure 3.1: Open Office setup in a library room in the USA

ability are well-documented [72] and have been shown to transfer to VR [102], we aimed to balance male and female participants we selected for our study. The participants had an average age of 27.6 years (± 6.7). Among participants, 44% reported currently working in an open office and 89% had little or no prior experience with VR.

3.2.2 Location and Technical Setup

The study was run in a total of three venues: two in the USA and one in Switzerland. In the USA, six sessions took place in a study room of a university library (Figure 3.1) and five sessions were run in a small conference room at a co-working office. In *Switzerland*, we ran twenty-four sessions in an office of a university (Figure 3.2). Between the three venues, the setup was similar: each room had windows and was well-lit. The participant, researcher, and assistant sat together at a large desk, with the researcher positioned near the participant, and the assistant positioned across. The main experimenter was present at all locations to ensure consistency (technical setup, room, etc.) and we carefully trained the assistants to perform distractions in the same way.

Technical Setup. For VR conditions, we used the HTC Vive Pro Eye with the Razer Blade 15 laptop with Nvidia RTX 2080 Max-Q graphics. An external screen was used to project Lightbot into the virtual desktop. For the non-VR conditions, only the laptop was used. Participants used a computer mouse to perform tasks, both in and out of VR. In VR, this meant that participants could not actually see the mouse and hand. Though this felt somewhat unnatural, in pilot testing we attempted using VR controllers in all conditions (with and

Figure 3.2: Closed Office setup in a university lab in Switzerland

without VR), and that proved even less natural. We also limited interaction to mouse-only, so there was no need to reacquire the mouse (e.g., from the keyboard) at any time.

Recordings We ran screen-recording software to document task completion in Lightbot and filmed the experiment with a GoPro camera in order to review interactions later.

3.2.3 Factor: Open vs. Closed Office

Simulating the open office. The simulation of an open office environment was one of our big puzzles, one which we considered at length. On the one hand, the conditions needed to be reproducible (across task sections and between participants). On the other hand, it had to be believable and thus involve live actors. The resulting setup consisted of scripted interactions between two actors with a backdrop of recorded office sounds, including snippets of conversation, typing, chewing, humming, rustling paper, and other office noises. Additionally, for a feeling of authenticity, one researcher and assistant simulated open office distractions. These included holding a simple conversation, occasionally getting up and walking to get something on the other side of the room, and other casual actions native to the office. Apart from a slight change in conversation (to avoid repetition), we practised the distractions to be believable and consistent between conditions and participants. We felt the result is as reasonable a simulation of an open office as could be achieved without recruiting more assistants. We include the study materials, including a copy of the audio track and VR scenes to the Open Science Framework for researchers wishing to reproduce our experiment [1].

[1]https://osf.io/ajx9s/

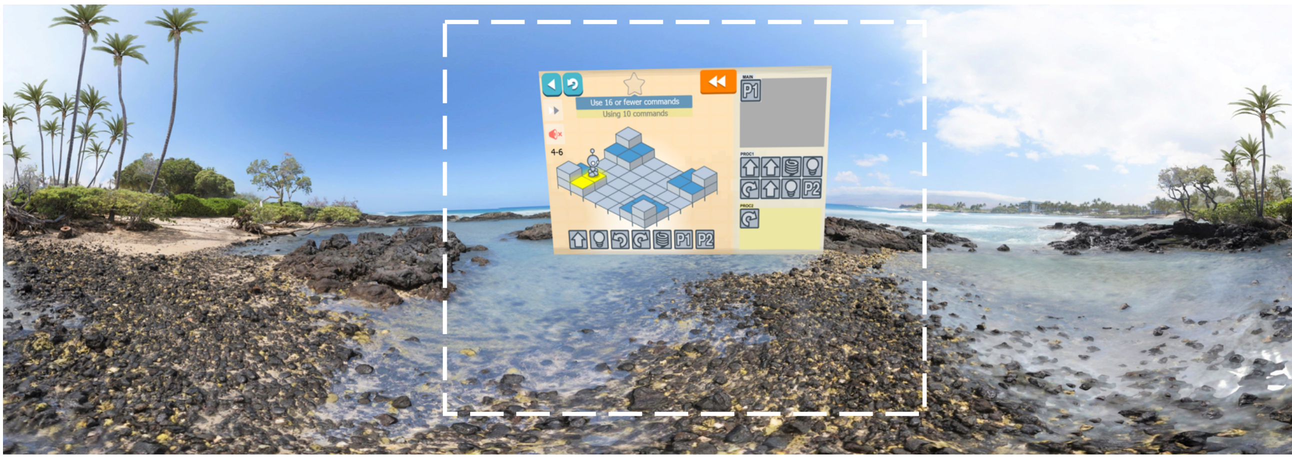


Figure 3.3: Virtual Hawaiian Beach, an environment created by Steam from photos of Big Island, HI[2]and shown here as a panorama.

Closed office. The closed office condition was simpler. The participant stayed in the same room, but the researcher and assistant stepped outside for twelve minutes while the participant performed tasks. The participant was instructed to start as soon as the door closes, and to stop when the door reopened on the researcher's return twelve minutes later. This way we modified the environment as little as possible, removing only the people and office noise.

3.2.4 Factor: VR vs. No VR

For all conditions, the participant was seated at a desk and performed their task using only the computer mouse. For the two non-VR conditions, participants viewed the task on a 15.6" laptop monitor (1080p). For the two VR conditions, participants put on the HTC Vive headset and were allowed a few moments to adjust the fit and focus. The desktop screen with the task was projected on a rectangular surface in the VR environment, so that the participants saw a floating screen before the presented VR background.

3.2.5 Conditions

We combined these factors into the following four conditions:

Closed office (with no VR): This served as the goal state for comparison against the other conditions. In this condition, the researcher and assistant left the room and gave participants twelve minutes to work through the tasks alone on the laptop (Figure 3.2).

Open office (with no VR): This served as a baseline to beat. The participant performed tasks on the laptop with the researcher and assistant seated at the same desk (Figure 3.1, without a headset). An audio track with open-office noises played in the background while the actors simulated office interactions.

Beach VR (open office with a beach VR environment): Same office setup as above. We used the free Hawaiian Beach environment from Steam VR (Figure 3.3), which features a photorealistic sandy beach, blue skies and palm trees, with wave animations and sounds. The task window was projected in a virtual desktop.

Office VR (open office with a simulated office VR environment): We used a 20-minute recording of a graduate student lab at the university, filmed using a stereo 360° camera. The virtual environment (Figure 3.4)

[2]For Figure 3.4 and Figure 3.3, the Lightbot overlay was not captured in the screenshots and was reintroduced manually to approximate what the user would have seen. The dotted rectangles indicate the user's approximate field of view.

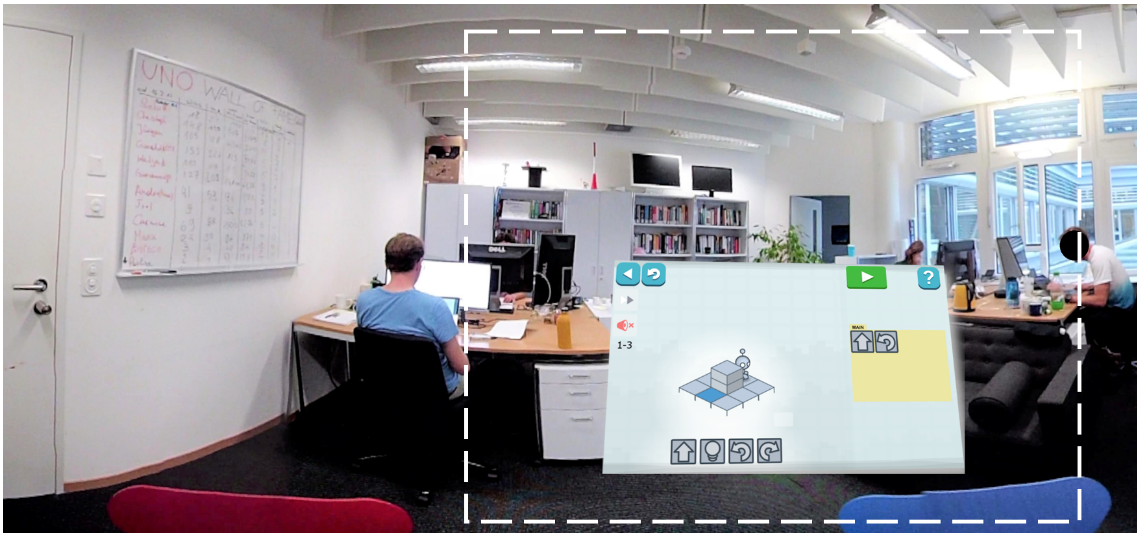


Figure 3.4: Virtual Open Office setting, 3D video filmed by the authors in an open office (note: cropped from full panorama).

features an office setting with four actors who work on computers at their desks and occasionally walk around, engage in conversation, and perform other actions similar to our live open office experience.

We used a within-participants design and counterbalanced these four conditions using a random Latin square.

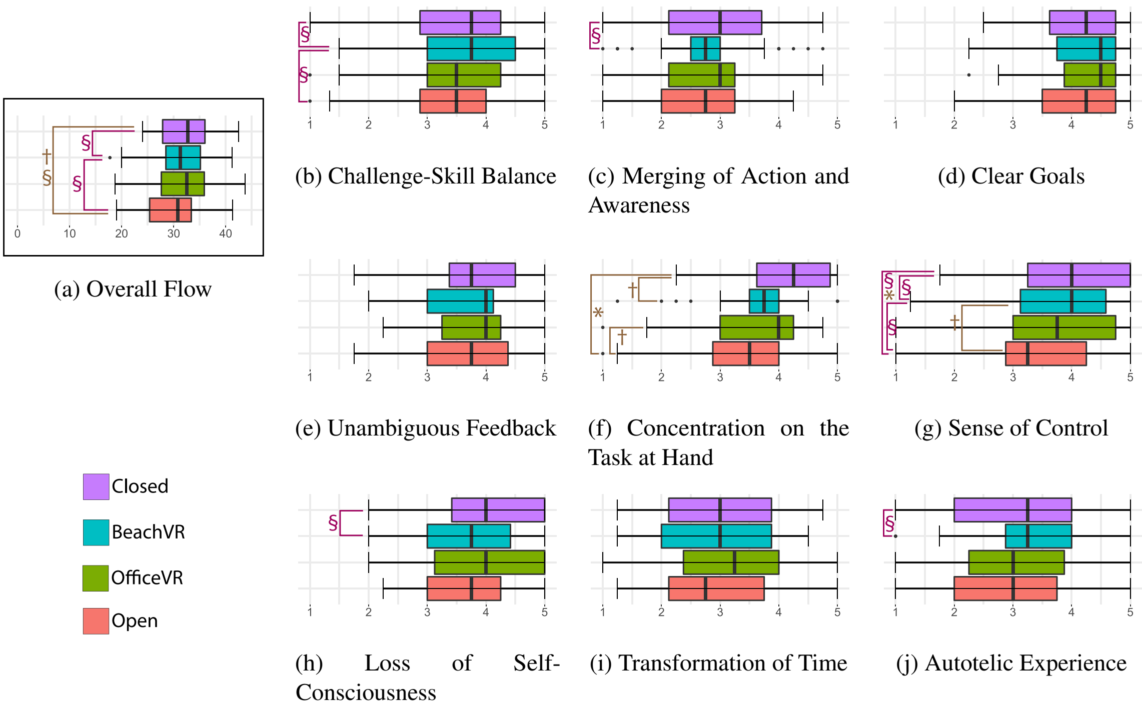
3.2.6 Task

For each condition, participants solved tasks in the coding game Lightbot [115], featured in Hour of Code. The fun, progressing challenges make the game a great candidate for inducing flow.The goal of the game is to navigate a little bot on a 2D board and light up all blue tiles. Rather than typing commands, the player adds a sequence of premade commands to an execution window by clicking on the appropriate tiles. The game gets progressively more challenging with a more complex board, new commands, and the introduction of procedures. With its simple set of commands, the game requires no prior experience and was thus accessible to a wide audience. At the same time, the programming nature of the tasks requires the same type of logical thinking that knowledge workers do on a daily basis. These aspects made Lightbot a good fit for our experiment. After extensive piloting of various tasks, we identified the LB tasks as a good choice that fulfills these criteria. Ideally, we would have chosen tasks with the same difficulty level to have full comparability. The variation of challenge in LB tasks made it harder to compare performance between conditions later on, as we will explain in the results. However, since the concept of flow requires the challenge of the task to match the increase in the participant’s skill over time, we identified the LB tasks as a good fit. In addition, the LB tasks assume no prior knowledge of programming, which made them accessible to a broad audience, and the increase in challenge of the LB tasks kept participants interested.

3.2.7 Procedure

Each study session lasted approximately two hours and had six stages: introduction, four conditions, and wrap-up.

Introduction. The participant was seated and given an overview of the study. They were then asked to



(a) Overall Flow
(b) Challenge-Skill Balance
(c) Merging of Action and Awareness
(d) Clear Goals
(e) Unambiguous Feedback
(f) Concentration on the Task at Hand
(g) Sense of Control
(h) Loss of Self-Consciousness
(i) Transformation of Time
(j) Autotelic Experience

Figure 3.5: Box plots of the main effects of environment on the overall flow scale and its nine dimensions. Post-hoc pairwise differences are indicated (†: $p < .08$, *: $p < .05$, **: $p < .01$) as well as pairs that had a significant interaction with order for some order of conditions (§).

complete the first six tasks of level one in Lightbot as a tutorial, in order to familiarize themselves with the program. At this stage, participants were allowed to ask questions if any confusion should occur. This setup had two variations in order to best match the tutorial to the participant's first condition. If the participant's first condition would be without VR, the participant performed the tutorial also without VR, simply using the laptop. If, however, the first condition was in VR, the participant performed the tutorial in VR. For this we used the Steam VR home "Summit Pavilion" environment, which we felt was a neutral option.

Four Conditions. Each participant was randomly assigned one of the four condition orders. For each condition, the participant started with task one of a new level (Level 2 for the first condition, Level 3 for the second, and so on). They would then be given 10-12 minutes to work through the tasks, until told to stop. In the end, we only counted tasks fully completed before the ten minute mark. However, we allowed participants up to two extra minutes to finish their current task, in order to reduce possible frustration and avoid negatively impacting the flow score. At the end of the condition, the participant was asked to fill out a 36-question survey on the laptop screen to measure their current flow state. Participants were invited to take short breaks if needed.

Wrap-up. Upon concluding the four conditions, the participant was interviewed one-on-one with the researcher for approximately ten minutes to gain an overall understanding of their experience in the experiment and their preference for the conditions. Finally, the participant filled out a demographics questionnaire and was paid.

3.2.8 Measures

We measured: flow, task completion, and condition preference.

Flow Scale. At the end of each condition, participants filled out a flow scale survey based on Jackson et al.'s Flow Scale Manual [61] to measure the participant's level of flow and engagement during the section. The survey measures flow based on nine major factors, e.g. the challenge-skill balance, loss of self-consciousness, and transformation of time. The questionnaire consisted of 36 statements which the participants rated on a scale of 1 (never) to 5 (always).

Task Completion. For each condition, we counted the number of tasks fully completed before the 10 minute mark. Participants were actually allowed up to 12 minutes to work through a level in case they were stuck on a task, in order to reduce potential frustration and affects on the flow measures. However, we only counted tasks fully completed after ten minutes.

Condition Preference. At the end of the study, participants were interviewed and asked to rank conditions in terms of how well they focused.

3.3 Results

Our observations suggested that, despite our efforts to use a known-to-be-effective progression of programming challenge, the order of Lightbot tasks influenced participants' experience of flow. We therefore performed an exploratory RM-ANOVA with the four environment conditions (closed office, beach VR, office VR, open office) as a within-participants factor that included the order of conditions as a between-participants factor. Our dependent measures were the nine dimensions of the flow scale, the overall flow score, and the number of Lightbot tasks completed within each condition. This analysis is consistent with the guidelines of the Flow Scales Manual [61], and we note that our choice of RM-ANOVA over non-parametric tests is an accepted practice [86]. Our post-hoc analyses used the Bonferroni correction.

3.3.1 Flow Scale

While we now know that task ordering influenced participants, for completeness we first analyze the results without considering task ordering. In this case there were significant main effects of environment for the following flow scale dimensions: concentration on the task ($F_{3,93} = 6.8$, $p < .001$, $\eta_p^2 = .18$), sense of control ($F_{3,93} = 4.9$, $p < .01$, $\eta_p^2 = .14$), and loss of self-consciousness ($F_{3,93} = 3.6$, $p = .02$, $\eta_p^2 = .10$). The main effect of overall flow was not significant ($F_{3,93} = 2.5$, $p = .06$, $\eta_p^2 = .08$).

Post-hoc tests (Figure 3.5) revealed that the closed office environment was rated higher than open office for the flow dimensions; this difference was significant for concentration on the task and sense of control ($p < .05$), and not significant for overall flow ($p = .06$). While not statistically significant, closed office was also rated higher for concentration on the task than both beach VR ($p = .06$) and office VR ($p = .06$) and higher for loss of self-consciousness than beach VR ($p = .06$). Interestingly, beach VR was also rated higher than open office for sense of control, but this was again not statistically significant ($p = .06$). Note that these differences, even though not accounting for ordering, show a difference between the closed and open office environments.

Order Effects

As mentioned, the LightBot tasks influenced participants and so we further analyzed our data with this in mind. While there were no main effects of order itself ($F_{3,31} < 2.4$, $p > .09$) there was a significant interaction between environment and order on overall flow score ($F_{9,93} = 2.8$, $p < .01$, $\eta_p^2 = .22$) and on the following dimensions: balance between challenge and skill ($F_{9,93} = 3.0$, $p < .01$, $\eta_p^2 = .23$), merging of action and awareness ($F_{9,93} = 2.2$, $p = .03$, $\eta_p^2 = .18$), sense of control ($F_{9,93} = 3.1$, $p < .01$, $\eta_p^2 = .23$), loss of self-consciousness ($F_{9,93} = 2.2$, $p = .03$, $\eta_p^2 = .18$), and autotelic experience ($F_{9,93} = 2.4$, $p = .02$, $\eta_p^2 = .19$).

Post-hoc analyses on these interactions revealed several differences for participants that saw conditions in the order: open office, office VR, beach VR, closed office. The pattern was that, in this condition, beach VR was rated significantly higher than closed office for balance between challenge and skill ($p = .01$), merging of action and awareness ($p = .04$), sense of control ($p = .03$), autotelic experience ($p < .01$), and overall flow ($p = .02$). Beach VR was also rated higher than open office for balance between challenge and skill ($p = .03$), sense of control ($p = .04$), and overall flow ($p = .01$). Beach VR was also rated higher than open office for balance between challenge and skill for participants that saw conditions in the order: beach VR, open office, closed office, office VR ($p = .03$). When accounting for ordering, beach VR appears to help users achieve flow.

For sense of control and overall flow, participants also rated the closed office condition higher than both beach VR (control: $p < .01$; flow: $p = .02$) and open office (control: $p < .001$; flow: $p = .02$) when they saw conditions in the order: closed office, beach VR, office VR, open office. Closed office was also rated higher than beach VR for loss of self-consciousness when participants saw conditions in the order: beach VR, open office, closed office, office VR ($p = .03$). As when not accounting for ordering, the closed office also appears to help users achieve flow for certain orderings.

Summary

Our findings show a tendency that participants were most in flow in the closed office environment, followed by the beach VR environment, and least in flow in both the office VR and open office environment without VR. However, while the closed office led to significantly higher flow than the open office, the other differences were either not significant or only true for specific orders of presentation of the conditions. While this tendency is not strong statistically, it is fairly consistent, and it is corroborated by our qualitative findings (see below). The qualitative findings also shed light on possible reasons for the lack of clearer differences.

3.3.2 Task Completion (Order Effects)

There were no main effects of environment ($F_{3,72} = 0.1$, $p = .94$, $\eta_p^2 = .01$) or order ($F_{1,24} = 1.2$, $p = .33$, $\eta_p^2 = .13$) on number of tasks completed. There was, however, a significant interaction between environment and order ($F_{9,72} = 9.9$, $p < .001$, $\eta_p^2 = .55$). Pairwise post-hoc analysis revealed that for every order, the final condition seen by participants always had the fewest tasks completed, and this difference was significant ($p < .05$) for all but three pairs.

Summary

It should be noted that Lightbot tasks are not designed to be equivalent in duration; however, the number of tasks completed can be considered an indicator of difficulty. This analysis therefore highlights that difficulty related more to the order of levels in Lightbot than to the conditions in our study. While on the one hand, this could be thought of as a confound in our study (flow depends on the level of challenge) and we encourage the reader to consider our quantitative findings to be exploratory, the fact that the flow dimension analysis did *not* have this same last-condition dominance lends more weight to the observed effects of environment on flow. Specifically, in the flow analysis, the environment condition accounted for more of the variance and participants consistently rated closed office highest, followed by beach VR, then office VR and open office; conversely, in the task completion analysis, the Lightbot levels were a better predictor of difficulty.

3.4 Interview Findings

Based on the transcripts of the semi-structured interviews with all participants, we derived participants' preference ranking for the four conditions and two authors performed a thematic analysis [20] and independently

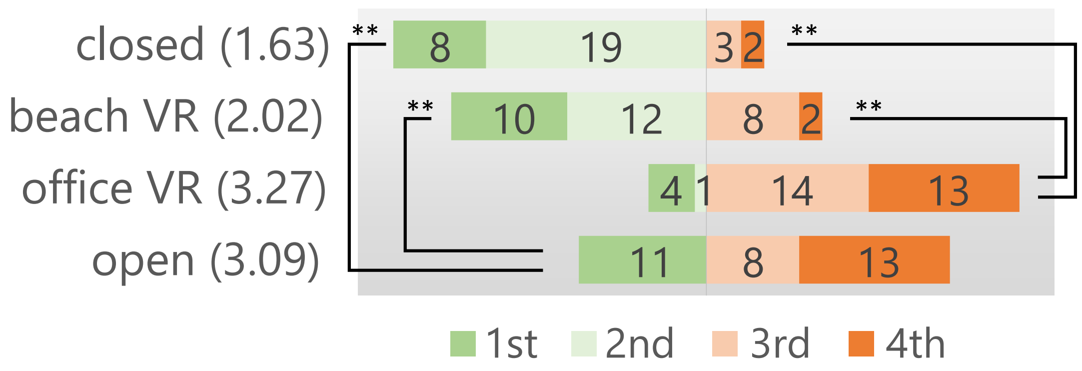


Figure 3.6: Rankings per condition (Mean Rank from Friedman's ANOVA indicated in brackets). Lines between conditions indicate significant pairwise differences (**$p < .001$).

coded the statements of all study participants. Both authors identified emerging themes in the data, and discussed and merged them iteratively. In the following, we report on the results and the main themes and observations.

3.4.1 Preferred Conditions

As part of their post-experiment interview, participants were asked to rank the four conditions according to preference. We received a complete ordered ranking for 30 participants, partial rankings for two participants who considered two conditions equal, and no ranking for one participant who felt that all conditions were equal. We performed a Friedman's ANOVA to compare ranks (Figure 3.6), which was significant ($\chi^2(3) = 37.5$, $p < .001$). Post-hoc tests revealed that both closed office and beach VR were ranked higher than office VR and open office environments (all pairwise differences $p < .001$). There was no significant difference between closed office and beach VR ($p = .27$) nor between open office and office VR ($p = .53$).

The high rankings of the closed office and the beach condition were supported by many positive comments. For instance, participants stated that the closed office "*was silent and I could focus the best*" *(Z14)* and "*it was very easy to stay focused*" *(S10)* in it, and for the beach VR condition that it was their "*favorite*" *(S12)* and that in "*the beach VR [they] could do a lot more, like fully concentrate*" *(Z21)*.

These positive comments contrast with the mainly negative comments from the open office conditions. For instance, for the open office without VR one participant stated that "*that was my least favorite because I could hear you clearly, and I didn't like it. As well, I was like... I was trying to focus, but then I was also listening to you too, so that's why I don't like it.*" *(Z05)*. Another participant elaborated, saying:

> "*I mean it's really night and day [between the closed and open office]. At some moment in time [in the open office setting], one of you shared a meme with the other person. Immediately, became more interested in that than what I was doing I was like 'I want to see too.' That's why I leave the office.. because.. I just need focus.*" *(S03)*

The lack of difference between the closed office and beach VR condition is also supported by the similarity in participants' comments about these two environments. For instance, participants commented for both conditions that it was easy to stay focused in them because "*there just weren't any distractions*" *(S10)* in the closed office and that they were "*not really aware of anything else in the beach*" *(S02)*. Similarly, the comments on the open office and open office VR conditions overlapped often, since they focused a lot on the experienced distractions, such as the background noise that "*was the worst*" *(Z09)* as someone stated or another one saying "*when there were background noise or talking, that's where my mind was sometimes* *going back and forth*" *(Z21)*. This contrast in experiencing distractions or not further illustrates the preference of the closed office and beach VR conditions to the open office and open office VR condition.

> The closed office and beach VR conditions were both significantly preferred over the open office and open office VR conditions.

3.4.2 Distractions

Participants mentioned an array of impediments to achieving focus and flow. In general, they differentiated between *auditory* distractions, such as people talking or chewing nearby, and *visual* distractions, such as people walking by. There are, however, also other more subtle sources for distractions, such as feeling of someone "*looking at you*" *(S02)* or "*too many things that [you] can do*" *(Z02)* in the place you are in. Overall, auditory distractions were mentioned the most and were in many cases also perceived as more distracting than visual ones: "*in the end, I will look at my screen and if I am focused, I will ignore what is happening outside of my screen so to say, but the noise is harder to cancel out I think.*" *(Z20)*. What was considered an auditory distraction depends on factors, such as the familiarity of the people talking or the content of the conversation:

"*It depends. If it's just background noise then it's fine. If it's really a conversation that I may be even a bit interested in, like your cat. I was like "what, someone is talking about cats?!". So I automatically pay attention to that so that's distracting. If it's just background noise, it's fine. So in terms of distraction, I would say interesting conversations are the worst distracting, [more] than movement and than background noise.*" *(Z03)*

At the same time, how distracting something is perceived is very individual and task dependent. While some prefer "*to be in a quiet environment alone*" *(Z19)*, for others it can be "*too quiet*" *(Z07)* or "*so silent [in the closed office that they] didn't like it very much*" *(Z06)*. Several participants stated that they like working in environments where there is white background noise, such as a coffee shop or library, or they listen to music, mostly without lyrics, such as "*soundtracks from video games*" *(Z13)*. Such constant noise can help to block out distractions: "*I usually always listen to music .. then I can't hear others making noise with their papers and stuff*" *(Z02)*. Similarly to the individual differences, the type of task has an influence on the perception of distractions and what works to block them out, as for instance one participant stated "*it really depends on the task. But I either seek out complete silent or coffee shop where there's like background noise*" *(Z09)*. Participants often mentioned adjusting the music and listening behavior to the task: "*When I study for an exam, I don't listen to music. But when I write a report for a laboratory, or stuff like that, then there's always music*" *(Z07)*. Finally, the perception of certain sources as distractions can also change over time and the tolerance threshold can increase:

"*I didn't use to, but I think I've gotten use to it. Now I can focus quite well. I still prefer no one around me. When I was at uni I would always go to the silent bits of the library, where no one is talking, and even the smallest distraction would really annoy me. But having work now, I think it's much easier to focus in an open office, just block it out.*" *(Z24)*

Furthermore, the effect of the distraction varies strongly, including people losing focus and concentration, their "*mind ... wandering off*" *(Z21)* and they switch away from their work task: "*I have very good peripheral vision. So, I have a hard time not looking up or leaning towards an object, therefore distracting myself*" *(S12)*. Participants also mentioned getting frustrated and annoyed, which in turn impedes them from getting into flow and decreases their focus at work.

> Distractions are often of auditory or visual nature and their perception and effect varies by individual and task.

3.4.3 (Beach) VR Effects

In our interviews, participants commented often on the effects of the VR, in particular the beach VR condition. Most participants noted that the beach VR helped them focus, to “*wash away*” *(Z07)* and reduce external interruptions or distractions: “*I was more focused with the beach I would say. I was able to focus ... I didn’t even notice you guys were there*” *(S03)*, or “*suddenly [in the closed office] I caught myself looking around a bit, and just looking at the office, and just getting annoyed by the computer sounds, like the air ventilation of the computer, and that was all gone in the beach setting.*” *(Z07)*. By placing users in a different environment, it also reduced the options for self-distraction that exist even in quiet rooms:

“*It would take me away from my reasons for not liking to study or work at home, because it is a home environment, and I always [think]... ‘I’m going to go to the bathroom. Oh, my bed is right there. I’m going to sleep,’ which is not a good thing. So, if I were to wear a headset, it would at least transport me away from my bedroom, even though I’m in my bedroom.*” *(S12)*

Additionally, participants valued the feelings and mood the beach environment induced, such as a feeling of calm and being relaxed and commented on the importance of the kind of VR environment used: “*the beach one. It just affected my mood, sitting around the beach, and some nature environment feeling. And then, this way, I could concentrate better. The VR environment is very important*” *(Z21)*. Several participants also explicitly mentioned the positive effect of the feeling of not being watched in the beach environment compared to the others: “*because I felt that I was [alone] and nobody was watching me and it was quite relaxing to see the beach.*” *(Z15)*. Finally, one participant also commented on the ability of the VR environment to foster creativity since in the right environment “*your brain starts to think in other way[s]*” *(Z23)*.

However, three participants also explicitly mentioned that the beach environment itself can be distracting, since “*you are not in a working environment, that makes it less focused*” *(Z14)* and the beach “*was just not the place to be on the computer*” *(Z17)*. A further negative effect mentioned with respect to the VR is the social detachment, since the VR “*detach[es you] so much from everybody else socially*” *(Z13)* while it is often good to have some people around for work for social reasons.

> VR can reduce distractions and increase well-being for many, but it matters a lot which VR environment is used.

3.4.4 VR Usage Experience

When asked whether they would consider using VR for short periods each day, many participants commented that they would for specific tasks, but that, especially due to its comfort, they wouldn’t wear it for all day and that it also depends on which other environments are available. One participant, for instance, mentioned

“*I mean, like for a few minutes, 20 minutes. That would be no problem, but I think if I had to wear this the whole day, it would be too heavy. If it would advance in a technical way, like weigh less and high resolution, I could imagine wearing it all day. Yeah, if it’s not physically a disturbance.*” *(Z20)*

Overall, many participants stated that the VR was comfortable enough for short periods of time, however, several issues were also raised about the experience wearing and using the VR and its comfort, in particular with respect to its weight, its heat, the blurriness at the edges, the eyes hurting after a while and for two participants also the dizziness they experienced in the VR condition.

> VR is an option for most but only for shorter time periods, especially due to comfort concerns.

3.4.5 Varying Task Difficulty

Independent of the conditions, we also asked participants about their experience performing the given tasks. While many enjoyed the tasks, stating, for instance, that they "*liked [them] quite much*" *(Z16)* and that "*it was fun*" *(S03)*, most participants commented on the varying level of difficulty of the tasks and that the later ones were more difficult. While some considered the increasing difficulty level a fun and engaging challenge, "*The beginning I felt like, a little more bored, but the later ones I was definitely very engaged. So the more challenging tests were more engaging.*" *(S10)*, others even perceived it as frustrating and annoying, for instance, "*I was into it at first, but then when I got to the ones that I couldn't do, I just was like, 'Let's just keep it going on repeat until the 12 minutes are over.'*" *(Z24)*, or "*Except for the last one, it was fun actually. I liked them. Just the last one was really frustrating.*" *(Z06)*.

> Our study tasks were generally well-suited for keeping participants engaged, however, participants perceived a significant increase in their level of difficulty.

3.5 Discussion

The goal of our study was to examine whether we can create a VR environment for an open office environment that provides similar benefits to that of a closed office for work. The qualitative and quantitative results of our study show that this idea has potential. Our qualitative results show that participants have a clear preference for the beach VR over the open office and open office VR conditions, as seen in the interviews from both countries. Our exploratory quantitative analysis provides evidence that further supports this observation, yet task ordering and task difficulty clearly had an effect on results. In the following, we will discuss the implications of our research.

3.5.1 Overcoming Workspace Limitations

Extensive planning goes into the creation of today's workplaces to best support knowledge workers. At the same time, companies have limited resources in terms of space and money, and despite the knowledge that open office workplaces can reduce workers' satisfaction and productivity, many companies have or are moving towards open office work environments [21]. One way that some companies are trying to overcome the increased number of distractions in an open office environment is by providing noise cancelling headphones or mounting white noise speakers in the office. Yet, since the distractions that knowledge workers experience in an open office are not just auditory, these solutions only partially address the problem. The results of our study show that we might be able to take advantage of VR, at least for certain periods of time, and "transport" the knowledge worker to a more desired work environment, regardless of physical space limitations. With the recent advances in VR technology there is a huge potential to better support knowledge workers in the future, especially where closed offices are not a viable solution. Even outside of the office, such as the home or the airport, where the distractions are of a different nature, we believe this approach could improve focus.

3.5.2 Tailoring the Environment to the Individual and Task

The data of our study shows that the details of an ideal work environment are dependent on both the individual and the task. While some participants prefer complete silence in a closed space, others prefer a coffee shop, or adjust the type of music they listen to based on the task they are working on. Similarly, while most participants in our study enjoyed working in the beach VR environment and several even preferred it to the closed office, some participants did not like the the mixing of work with a beach environment which they considered more for leisure. Fortunately, VR is highly customizable and one can think up an infinite number of environments to work in. Users of a VR office could tailor the work setting to their needs and preferences, even changing environments to suit their task. For instance, when knowledge workers have to perform rote work, they can choose a more stimulating environment, for cognitively demanding assignments possibly a more quiet and calm environment, and for creative tasks an environment that fosters creativity, e.g. by increasing the blue light portion [1]. In the future, we might also be able to automatically adjust the environment to the individual on a moment-by-moment basis by using biometrics (e.g., [99]).

3.5.3 One Environment Does Not Fit All

While VR has great potential to help knowledge workers achieve flow, we are not suggesting that they work in VR all day long. Today’s headsets have clear limitations, especially in terms of comfort, screen resolution, and weight. Additionally, the headset can lead to social detachment, which participants mentioned several times, and causes challenges in interacting with the real world. For instance, based on our experience using the VR extensively for this study, drinking coffee while working in VR is a challenge at best, if not outright dangerous. Therefore, the goal is not to completely replace the usual work environment with VR environments, but to provide the opportunity for transforming the work setting to achieve the right state of mind. Future studies should examine for which situations a VR environment is best suited and how well it can integrate into the usual day of a knowledge worker’s life. Furthermore, we must consider the macro effects that such technology would impose on the workplace. Though virtual reality could allow participants to escape to an isolated world and fully focus on their work, it would be at a trade-off with the collaborative factors of the open office. It would be interesting to study how VR technology would play out on the social dynamics of the workplace.

3.5.4 Choosing Study Tasks

The goal for our study was to examine the effect of different environments while performing work tasks. Therefore, we tried to compile study tasks that met many constraints. They had to be engaging and challenging, yet not too difficult; flow inducing, but not too time-consuming; comparable with each other, yet doable by a broad audience. At the same time, these tasks had to comply within current VR technology and its limitations, such as the limited resolution of headsets and the difficulty of VR keyboard input. After several iterations and extensive piloting, we chose the Lightbot tasks, programming tasks that can be solved by people even without programming experience. However, despite the program’s organization into levels of increasing complexity which slowly builds upon previous levels, users felt that the difficulty varied dramatically for certain levels, as their comments have shown. In general, it is challenging to identify a good set of comparable work tasks that fit the requirements of such studies. Identifying and sharing such tasks with the research community could significantly facilitate research in this domain and its generalizability. Thus, while our quantitative results are less generalizable due to the ordering effects, we believe that they provide initial evidence that illustrates the potential of VR work environments. Further studies are needed, either with a more comparable set of study tasks or a large number of participants to examine the generalizability.

3.5.5 Our Mixed-Methods Approach

Given the varying difficulty of the study tasks, we designed the study to use a mixed methods approach to evaluate the viability of using VR in realistic work settings. As with any study, there are tradeoffs that come with study design choices. In our study, we traded off some control on exact task difficulty and its comparability for the realism of an office setting with realistic programming-like tasks. By using a mixed methods approach, we were able to triangulate the qualitative with the quantitative findings. Note that our claims stem from a mixed methods approach and are not meant to reflect a more traditional hypothesis test (where the importance of p-values are paramount), but instead we take the approach of reporting exact p-values, eta squared effect sizes, using different symbols for different thresholds in our figures, and considering these all in the context of qualitative findings from the interviews with participants. We believe that this triangulation and the qualitative analysis is actually one of the strengths of our research and it is an important step in a larger body of work that can further triangulate these findings, perhaps with a larger sample or a longitudinal study.

3.6 Threats and Limitations

One threat to the validity of our study is the short duration of conditions. To avoid possible exhaustion, we designed our study to fit in two hours. With the time required for setup, transitions, and wrap-up, this gave participants only twelve minutes for Lightbot in each of the four conditions. Whether this is truly enough time to get into the flow state is hard to say, and is highly variable by individual. This may be the reason why we did not see significant differences in flow measures between conditions.

Furthermore, our simulated open office environment and the Lightbot tasks does not claim to accurately represent the daily environment and tasks of knowledge workers. To generalize our findings to the industry we would need to run longitudinal studies with knowledge workers in a real open office.

Some extraneous variables were particularly problematic in our experiment. Most of our participants were first-time VR users, and the novelty and excitement of using the VR may potentially affect the results more than our independent variables. Furthermore, the challenge of certain Lightbot levels (particularly in conditions two and four) required significantly more thought and time for completing tasks. As a result, performance was effected more by the task rather than by the condition. Although we tried to control for novelty and level difficulty through order randomization, the effect of these variables was so diverse between participants, that it introduced high noise levels. A longer-term field study would likely reduce these limitations. Another extraneous variable is the auditory factor, which is a natural part of the VR environment. Both VR environments included sound (beach waves or office noises) for an immersive experience. It would be interesting to explore how much participants are affected by the auditory factors as opposed to the visual.

Chapter 4

Productivity and Wellbeing in Teams

4.1 Introduction

The modern workplace demands more of its employees than ever before as technology assumes a larger role. Workers have to perform a broad variety of complex tasks, working in a global and fast-paced environment, and experiencing interruptions and distractions. In this context, the productivity and wellbeing of the individual as well as the team and the overall organizations is increasingly important, yet little is known about them and their relation. The broader vision of this part of our research is to better understand productivity and wellbeing on an individual and team level, and to develop approaches that support professionals in spending their time well at work.

There is an abundance of research on developer productivity examining aspects such as the quantification of productivity (e.g., [25, 34, 83]), developers' perception of productivity (e.g., [81, 82]), factors of developer productivity (e.g., [46, 84, 110]), and most recently the effect of the Pandemic on developer productivity [90]. There is also some research that investigated team productivity of software developers and how factors such as team cohesion or team size affect productivity (e.g., [70, 97]). The majority of this research examined individual productivity and team productivity separately, and we have yet to understand the correlation between them as well as their relation to wellbeing.

The objective of this part of our research is to address this gap and to better understand the different levels of productivity—-team and individual–their interplay and their relation to wellbeing in the workplace. Primarily, we are interested in the following research questions:

RQ1: What is the correlation between perceived individual and team productivity?
The answer to this question could play out in several ways. At first sight, it may seem most logical that as individuals are more productive, they see their team as more productive, because they, themselves are part of the team. However, it could also be that when individuals engage in more team-oriented tasks, for example assisting colleagues or sharing knowledge, they view their own productivity as diminished in favor of a heightened team productivity. Examining this question will allow us to see how developers weigh their contributions in regards to the team, and what dynamic plays out between the individual and team.

RQ2: What is the relation between team productivity and "softer" factors?
These softer factors can include experienced stress on an individual and team level, frequency and nature of unplanned work, interruptions, and factors that participants report in the open-ended questions. Does stress negatively impact productivity, or does it drive participants to be more productive? These questions will also help us gain insight into the broader research topic: the relationship between productivity and wellbeing.

RQ3: What is the relationship between social interactions and individual/team productivity at work?
The knowledge workplace is often a very collaborative environment. How does the frequency of interaction with teammates and supervisors affect productivity on an individual and team level? And does the nature of this interaction–task-focused or off-task–result in different effects?

RQ4: How do social interactions affect stress at work?
In a similar question, we explore how these social interactions impact stress. Is talking to a manager more stressful than talking to a teammate, or does it have a reassuring effect? Does the subject of conversation (on task or off) greatly change the effect on stress, or is it the frequency of interaction which counts? These answers will likely be closely related to the general dynamic within the team.

As part of our research for this thesis, we designed a long-term exploratory field study, developed tool support for collecting the relevant data and piloted the study with five software developers. Our initial insights from the pilot study show that the study design works well to collect the relevant data, yet that the concept of a "team" is not as well defined in the workplace as we assumed and that there can be little awareness of the team members, especially during the Pandemic. Based on our insights we made minor adjustments to the study design and are now running it with more than 70 participants.

4.2 Study Design

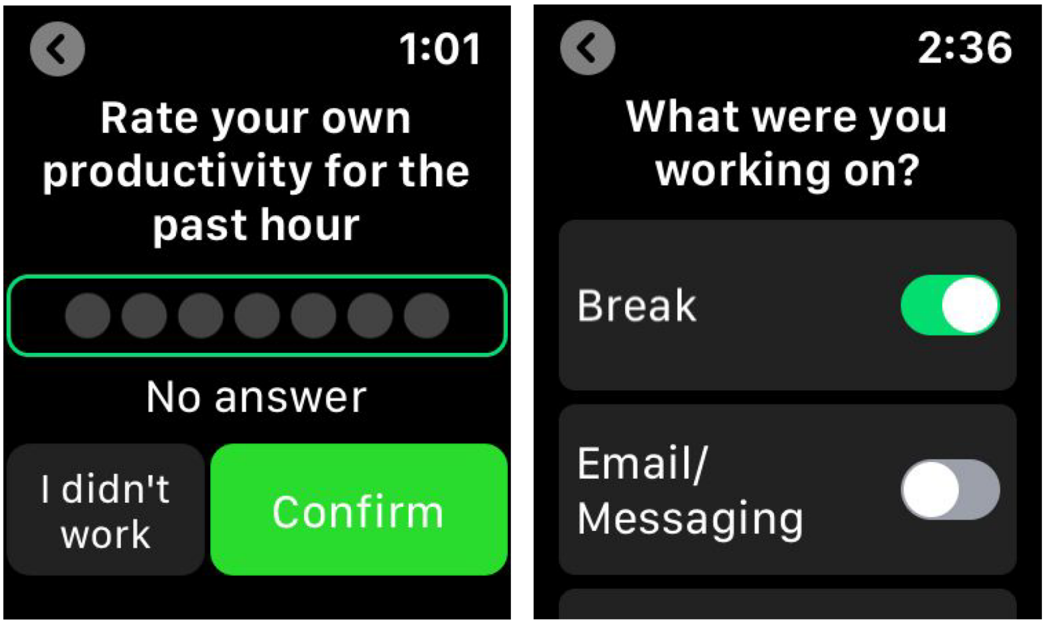
To address our research questions we designed a 6 to 8 week field study. For our field study we adopted a mixed-methods approach to capture both finer-grained and longer-term productivity patterns, collecting data on productivity–team and individual level–and on wellbeing. Specifically, we designed and developed approaches to collect fine-grained experience samples on productivity and wellbeing, daily and weekly diaries of work and productivity patterns, interviews to gather more in-depth qualitative data from participants on the study period, and a survey to collect demographic and personality data.

As an inclusion criteria for our study we require that at least 40% of a team participate in the study to obtain a representative sample. Furthermore, to ensure that our Apple watch and the Apple watch application we developed work properly, we require participants to have an iPhone 6 or higher.

Self-reports (every 2 hours)
The participant is prompted to respond to a self-report every two hours throughout their workday. The workday hours are defined by the user before the start of the study, and can be adjusted at any time on the watch. The self-reports pop up one hour into the work day and continue every two hours until the end of the workday, which typically results in 4 times per day. The self-report contains 3 Likert-scale questions (rated 1-7) and one activity choice question:

1. Rate your own productivity for the past hour (1-7) (see Figure 4.1a)
2. Rate your team's productivity for the past hour (1-7)
3. Rate your stress level for the past hour (1-7)
4. What were you doing? Select all the apply: Break, Email/Messaging, Meeting, Planning, Software Development Work, Browsing, Read/Write Documents, Other Individual Tasks, Other Team Tasks (see Figure 4.1b)

This Experience Sampling Method (ESM) style questionnaire is designed to capture productivity and stress patterns in the participant's day. By sampling experience, we are able to build an image of the participant's day instantaneously, rather than retrospectively. The advantage of quick and frequent sampling is that participants are more likely to remember their perceptions at that point in time. The Experience Sampling



(a) Productivity self-reporting

(b) Activity self-reporting (partial)

Figure 4.1: Screenshots of self-reporting in our Apple Watch application.

Method (ESM) has been used extensively in past studies and has been shown to have internal [55] and external validity [29]. This approach has been effectively used in several studies on developer productivity [50, 76, 80].

To reduce the invasiveness of our ESM approach and foster a high number of in the moment self-reports, we designed and developed study support in the form of an Apple Watch application that allows to answer the self-reports on the watch (see Figure 4.1). For our study, we plan to provide an Apple Watch to each study participant.

End-of-day and End-of-week Questionnaires

To complement the short self-reports with more thorough reflections, participants will be asked to fill out a longer (10 minute) questionnaire at the end of each day and week. This questionnaire is a diary-style survey which asks participants to reflect about the factors that affected their productivity and wellbeing. These surveys create a space to examine a much greater array of factors than the self-reports (for example, social interactions, unplanned work) and invite the participant to answer an open-ended question of what affected their productivity that day or week, allowing us to discover factors we previously may have overlooked. For the survey we developed a web application that allows us to collect the data from participants as well as it provides a visualization of the self-report data that participants entered on their Apple watch. For the study, we will remind participants about these surveys using a daily email reminder that is automatically sent by the web application. A screenshot of part of the end-of-day survey is illustrated in Figure 4.2).

Interviews

We will hold a series of three one-on-one video calls between each participant and the researcher: at the start, middle, and end of the study. In these interviews we ask participants about their work and team, and about their perspectives on productivity. These questions allow us to better understand what is usual for the participant (for example, how much time in a week is spent in meetings), as well as the dynamic within the team. In our diverse range of participating teams, we have found that some work very closely on tasks, while others do not share project work at all, but come together to discuss progress and share tribal knowledge.

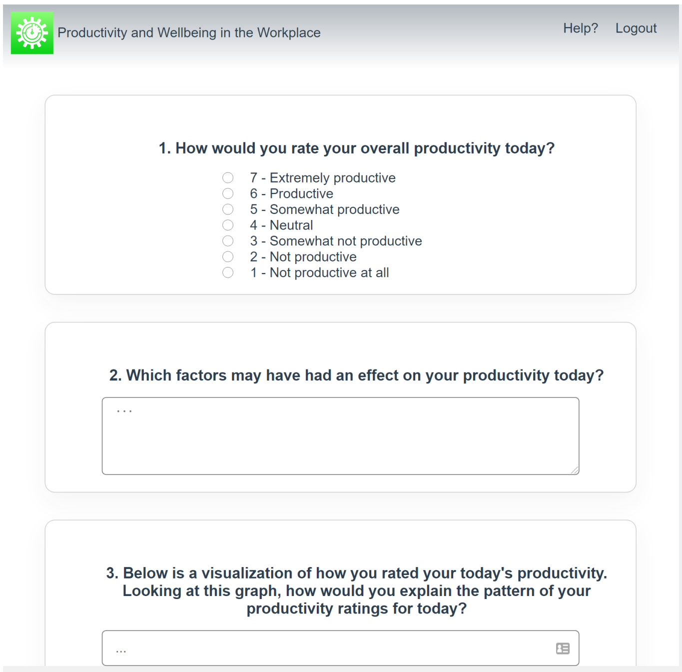


Figure 4.2: Screenshot of end-of-day online survey.

Final Survey
After the completion of the study, we will send participants a final online questionnaire. This survey serves to collect the participant's demographic data, and asks a set of 50 personality questions in order to determine the participant's personality type. These questions were previously used in several studies, including one by Grant et al., which explored implications of leadership and proactivity in teams [45]. We will relate these personality types to effects of different factors (such as interactions) and stress on individual productivity.

4.3 Data Analysis and Collection

Below we share a few notes of how we plan to answer our posed research questions with the data collected in the self-reports, surveys and interviews.

RQ1: What is the correlation between perceived individual and team productivity?
In the hourly questionnaires, participants are asked to assess their own productivity and their team's productivity for the last hour. We likewise include the same questions, but on a daily and weekly basis - in order to get a more coarse-grained time scope. By comparing the two productivity scores (normalized by individual) and using correlation analysis, we can explore how a change in perceived individual productivity correlates with perceived team productivity.

RQ2: What is the relation between team productivity and "softer" factors?
Participants report perceived productivity and stress on an hourly basis, and also on a daily and weekly basis. We question participants further on selected soft factors with several questions in the daily and weekly surveys, including levels of interactions, interruptions and unplanned work. For our analysis, we plan to examine whether changes in the softer factors correspond to changes in productivity levels. Additionally, we

plan to perform this analysis within teams (considering each team as a data point), to see if, for example, teams that have a lower stress level or lower/higher satisfaction are overall more/less productive.

RQ3: What is the relationship between social interactions and individual/team productivity at work? We ask participants to report frequency of team and manager interaction on a daily basis, and how related those interactions are to work. By correlating this data with perceived individual or team productivity, we can explore whether interactions positively or negatively impact an individual's productivity and the team's productivity as a whole. Similarly, we will analyze how this this moderated by the extent to which the interactions are task-oriented or relationship-oriented. Finally, we are interested to investigate if the impact of interactions with the team and the manager produce different results.

RQ4: How do social interactions affect stress at work? For this research question we can use the interaction data (frequency and work-relatedness) for the manager and team in conjunction with the collected individual stress data, averaged over the team. Using the collected data we can perform a comparison and examine whether off-task interactions with the team result in reduced stress (based on a correlation analysis), while work-related interactions with the manager increased stress, or if the situation is different than one might assume.

4.4 Pilot Study

To inspect the design of our study and test the study support we developed, we performed a pilot study. For our pilot study we recruited six professional software developers through research contacts. In the end, five of these six professionals participated in the study.

Throughout and at the end of the pilot, we examined the collected data and the participation. Overall, we successfully ran the study with the 5 participants for a total of 2 months and collected more than 2200 data points from the self-reports alone. The pilot showed that the overall design works well to collect the relevant data. At the same time, we made several interesting preliminary observations.

Productivity
As part of our pilot, we asked participants how they would define productivity. The most common answer was a variant of "output over input"–the traditional definition by [88] we discussed in Chapter 2. Yet, several participants said they also found activities such as brainstorming about a problem or helping a colleague to be productive, even when no direct output was produced. This supports the idea that measuring developer productivity requires a more nuanced approach. We expected divergent views of "productivity", since it has been contested in past research, yet surprisingly we were forced to question several other definitions we took for granted.

Definition of Team
In reporting team productivity, the concept of "team" turned out to be far from trivial. We have seen (in the pilot team and the variety of other teams we have recruited for the study), that this definition can be unique to every participant, and can in fact change regularly based on the task and context. One group we recruited after the pilot sees itself as a team because all members have the same position (sales executives) and share the same overall goal (to bring revenue). They come together every week to discuss problems and exchange ideas for solutions. And yet, individually each participant is assigned their own set of projects, on which they work in conjunction with other professionals of the company (e.g. a market analyst, designer, etc.). In the morning they may work with one project, in the afternoon with another. To accommodate this team fluidity, we explicitly added a statement on the concept of team and encourage participants to think of their team as the people they regularly interact with and with whom they share responsibilities for common outcomes, in the time span they are questioned about.

Full Study

Based on the successful pilot, we have now extended this study to a larger set of participants. We are currently running this study with more than 70 knowledge workers of 15 teams (though, again, this team concept can differ from the perceived team on which participants reflect). The participating teams originate from various countries including Switzerland, Germany, USA and Canada. In the set of participants, several knowledge work domains are represented, including software developers, executive accountants, insurance workers, and UX designers, to name a few. This study does not aim to be representative of all knowledge workers–there are too many domains to consider for this scope–but rather to gain preliminary ideas of how productivity and wellbeing plays out in a knowledge workplace. Before starting, participants are, of course, assured that their reports will not be shared with other members of their team, managers, or organization. Given the very different dynamics and objectives these various teams have, we will need to ensure to consider this in our future analysis.

4.5 Threats to Validity

The power of self-reflection. Past studies have found that the very process of reflecting on one's own productivity may in itself cause software engineers tot be more productive [17]. We mitigate this effect by disregarding the first week of data (which may have a learning/novelty effect). Furthermore, we are not interested in the absolute productivity ratings, but rather the change in ratings depending on progressing time and shifting factors. As a result, any learning or self-awareness effect should not have a major impact.

Pressure to overvalue. Participants may be tempted to overvalue their own productivity, especially if there is a chance their data will be viewed by management. We assure our participants that there data will always be collected and stored anonymously, and that at no point will it be shared with any individual (except, perhaps, themselves), and especially not with colleagues or management. If despite this, participants feel an unconscious pressure and inflate their productivity, this should not greatly impact our analysis. Again, we are not measuring the absolute, but rather the changes in productivity. Consistent overvaluing should not, therefore, have a significant effect.

Interpreting the scale. Participants rate on a Likert scale of 1-7. But what if your 6 is my 4? Again, this should not matter as argued above - we are not measuring absolute values, but rather changes. When analyzing the effect a factor (e.g. stress) has on someone's productivity, we will first normalize both values to see how these compare to the individual's average.

Chapter 5

Future Directions

When it comes to a more holistic notion of developer productivity, there is yet much to explore. It would be especially interesting to continue the thoughts of our two previous studies.

Boosting Flow through Virtual Reality. In our first study on the use of VR in team environments, we found that virtual reality can be effective in removing distractions and fostering flow, and that this technology has the potential to bring a noisy open office closer to the concentration capacity of the closed office. This research can be extended in several interesting ways. Our study featured a virtual beach environment as a first step, yet other environments might have even more positive effects. There is a wealth of research on the potentially calming effects of nature environments [7] or color spectrum [35, 69, 112, 118], which would be interesting to apply here. Perhaps the environment can even measure a person's current activity level or arousal and adjust the environment accordingly. In future work, we plan to examine an array of environments and their effect on stress, productivity and other aspects, such as creativity and fatigue.

Furthermore, it would be interesting to reintroduce some of the actual environment into the virtual. One problem of the current solution is that the user loses all context of their surroundings: the desk, the chair, the office. An interesting approach would be to map the actual space into the virtual (similar to a mixed reality situation), so that the user has some ability to navigate and orient in the actual world while reaping the flow-inducing and stress-lowering benefits of the virtual world.

Productivity and Wellbeing in Teams. Several research extensions naturally flow from our second study on the individual and team level productivity and wellbeing. Based on our initial insights, an interesting avenue would be to explore the design space for an approach on team awareness and examine its effects on individual and team dynamics. There are already a few research approaches that have been proposed in the past to provide more awareness for software development teams, such as FASTDAsh [18] and WIPDash [62], yet these approaches are mostly focused on work and do not take into account the social and wellbeing component. Especially with the current distributed working situation (due to the pandemic), an approach that provides a more holistic team awareness could be valuable not only in improving productivity, but also wellbeing.

It would also be very interesting to replicate our second study once the pandemic has slowed and developers have reentered the offices. Of course, the work situation will likely be somewhat different than it was pre-Corona. After all, such a monumental shift will have lingering effects – hopefully some of them positive, such as a greater emphasis on employee wellbeing. Still, it would be interesting to compare the dynamic of the individual and team between the normal office and distributed home working environments.

Chapter 6

Conclusion

In a set of two research studies, we explore the productivity and wellbeing of software developers in team dynamics. In study 1, we examined the effects of virtual and traditional work environments on knowledge workers while performing work-related tasks. The qualitative and exploratory quantitative results show that the closed office and the beach VR are similarly good according to users' ranking in reducing distractions and inducing flow, and that these two environments are preferred over the non-VR open office and VR open office environments. Further studies need to be run to examine the generalizability of our quantitative results.

Overall, these results indicate the potential that VR environments have to help knowledge workers achieve flow and stay calm and focused even in loud open office work settings. At the same time, the results open up new opportunities for research. Since the preferred work environment is highly individual, we need to examine which environment is best suited for which situation. Considering the high customizability of VR and the nearly infinite possibilities for environments, we may be able to tailor the VR work setting to individual preferences, the task, and possibly even the current mental state of the knowledge worker to provide the best experience in every moment. At the same time, we need to explore how this technology will alter workplace dynamics and social interactions among knowledge workers of the future, due to the current limitations of the technology.

In the second study, we designed and piloted a study in which participants regularly respond to a series of productivity and wellbeing questionnaires throughout their workday. The multi-tiered data collection includes hourly, daily, weekly, monthly, and one-time surveys for a better understanding of the fine- and coarse-grained factors which affect the productivity and wellbeing of software teams. We are especially interested in the correlation between individual and team perceptions of productivity and wellbeing, and what this means for the team dynamic. Having piloted the study, we are running it in full with a running count of 80 participants across 15 teams, from Switzerland, Germany, the USA and Canada.

In our future research, we aspire to innovate the modern approach to developer productivity. Though productivity is often thought of as work over time, research has shown that assessing developer productivity is complex. Furthermore, boosting developer productivity is a delicate task. Commonly, what starts off as a well-meaning effort to improve productivity turns into a rush to speed up the wheels and crank the most out of employees. This may result in short-term success, but in the end this approach is doomed to fail: a purely output-driven aim will exhaust many developers, possibly leading to burnout and other serious issues. Instead, we aim to get the golden egg without killing the goose. By taking a more holistic approach and incorporating developer wellbeing, we can design much more effective strategies. Work satisfaction and positive mood has already been shown to boost productivity, and productivity in turn can foster job satisfaction – an upwards cycle of productivity and wellbeing. In our future endeavours, we aim to further our understanding of developer productivity and wellbeing, especially in relation to individual and team dynamics.

Bibliography

[1] ABDULLAH, S., CZERWINSKI, M., MARK, G., AND JOHNS, P. Shining (blue) light on creative ability. In *UbiComp '16 Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (New York, NY, USA, 2016), ACM, pp. 793–804.

[2] ALBRECHT, A. J. Measuring application development productivity. In *Proceedings of IBM Applications Development Symposium* (Monterey, 1979), p. 83.

[3] ALLEN, D. *Getting Things Done : The Art of Stress-free Productivity*. Viking, New York, 2001.

[4] ALTMANN, E. M., AND TRAFTON, J. G. Memory for goals: an activation-based model. *Cognitive Science 26*, 1 (Feb 2002), 39–83.

[5] ALTMANN, E. M., TRAFTON, J. G., AND HAMBRICK, D. Z. Momentary interruptions can derail the train of thought. *Journal of Experimental Psychology: General 143*, 1 (2014), 215–226.

[6] ANDUJAR, C., AND ARGELAGUET, F. Virtual pads: Decoupling motor space and visual space for flexible manipulation of 2D windows within VEs. In *2007 IEEE Symposium on 3D User Interfaces* (2007), IEEE, pp. 99–106.

[7] ANNERSTEDT, M., JÖNSSON, P., WALLERGÅRD, M., JOHANSSON, G., KARLSON, B., GRAHN, P., HANSEN, Å. M., AND WÄHRBORG, P. Inducing physiological stress recovery with sounds of nature in a virtual reality forest—results from a pilot study. *Physiology & behavior 118* (2013), 240–250.

[8] ARI. Worry, gratitude & boredom: As covid-19 affects mental, financial health, who fares better; who is worse? *http://angusreid.org/covid19-mental-health/* (2020).

[9] BAILEY, B. P., AND KONSTAN, J. A. On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Computers in Human Behavior 22*, 4 (Jul 2006), 685–708.

[10] BAILEY, D., AND KURLAND, N. A review of telework research: Findings, new directions, and lessons for the study of modern work. *Journal of Organizational Behavior 23* (06 2002), 383 – 400.

[11] BAKER, E., AVERY, G., AND CRAWFORD, J. Satisfaction and perceived productivity when professionals work from home. *Res. Pract. Human Resour. Manage. 15* (01 2007), 37–62.

[12] BAKKER, A. B., AND BAL, M. P. Weekly work engagement and performance: A study among starting teachers. *Journal of occupational and organizational psychology 83*, 1 (2010), 189–206.

[13] BAO, L., LI, T., XIA, X., ZHU, K., LI, H., AND YANG, X. How does working from home affect developer productivity? – a case study of baidu during covid-19 pandemic.

[14] BARUCH, Y. Self performance appraisal vs direct-manager appraisal: A case of congruence. *Journal of Managerial Psychology 11* (1996), 50–65.

[15] BARUCH, Y. Teleworking benefits and pitfalls as perceived by professionals. *New Technology, Work and Employment 15* (12 2002), 34 – 49.

[16] BEECHAM, S., BADDOO, N., HALL, T., ROBINSON, H., AND SHARP, H. Motivation in software engineering: A systematic literature review. *Information and Software Technology 50* (08 2008), 860–878.

[17] BELLER, M., GOUSIOS, G., PANICHELLA, A., PROKSCH, S., AMANN, S., AND ZAIDMAN, A. Developer testing in the ide: Patterns, beliefs, and behavior. *IEEE Transactions on Software Engineering 45*, 03 (mar 2019), 261–284.

[18] BIEHL, J., CZERWINSKI, M., SMITH, G., ROBERTSON, G., AND BAILEY, B. Fastdash: A visual dashboard for fostering awareness in software teams. In *CHI 2007 Conference on Human Factors in Computing Systems* (January 2007).

[19] BILLINGHURST, M., AND STARNER, T. Wearable devices: New ways to manage information. In *Computer* (New York, NY, USA, Jan 1999), vol. 32, IEEE, pp. 57–64.

[20] BRAUN, V., AND CLARKE, V. Using thematic analysis in psychology. *Qualitative research in psychology 3* (Jan 2006), 77–101.

[21] BRENNAN, A., CHUGH, J. S., AND KLINE, T. Traditional versus open office design: A longitudinal field study. *Environment and Behavior 34*, 3 (2002), 279–299.

[22] BRIEF, A. P., AND ALDAG, R. J. Employee reactions to job characteristics: A constructive replication. *Journal of Applied Psychology 60*, 2 (1975), 182–186.

[23] BROEREN, J., RYDMARK, M., AND SUNNERHAGEN, K. S. Virtual reality and haptics as a training device for movement rehabilitation after stroke: a single-case study. *Archives of physical medicine and rehabilitation 85*, 8 (2004), 1247–1250.

[24] CASCIO, W. Managing a virtual workplace. *Academy of Management Perspectives 14* (08 2000).

[25] CATALDO, M., HERBSLEB, J. D., AND CARLEY, K. M. Socio-technical congruence: A framework for assessing the impact of technical and work dependencies on software development productivity. In *Proceedings of the Second ACM-IEEE International Symposium on Empirical Software Engineering and Measurement* (New York, NY, USA, 2008), ESEM ’08, Association for Computing Machinery, p. 2–11.

[26] CHRYSTAL, A., AND MIZEN, P. Goodhart’s law: Its origins, meaning and implications for monetary policy. *Cent Bank Monet Theory Pract Essays Honour Charles Goodhart 1* (01 2003).

[27] CLEMENTS-CROOME, D., AND KALUARACHCHI, Y. *Assessment and measurement of productivity*. 01 2000, pp. 129–166.

[28] CSIKSZENTMIHALYI, M., AND CSIKSZENTMIHALYI, I. Introduction to part iv. In *Optimal Experience: Psychological Studies of Flow in Consciousness*, M. Csikszentmihalyi and I. S. Csikszentmihalyi, Eds. Cambridge University Press, 1988, pp. 251–265.

[29] CSIKSZENTMIHALYI, M., AND LARSON, R. Validity and reliability of the experience-sampling method. In *Flow and the foundations of positive psychology*. Springer, 2014, pp. 35–54.

[30] CSIKSZENTMIHALYI, M., AND LEFEVRE, J. Optimal experience in work and leisure. *Journal of Personality and Social Psychology 56*, 5 (1989), 815–822.

[31] CZIKSZENTMIHALYI, M. *Flow: The psychology of optimal experience*. New York: Harper & Row, 1990.

[32] DAVENPORT, T. H. Thinking for a living: How to get better performance and results from knowledge workers. *Notes* (2005), 599–603.

[33] DAVENPORT, T. H., AND PEARLSON, K. E. Two cheers for the virtual office. *Sloan Management Review 39* (1998), 51–65.

[34] DEVANBU, P., KARSTU, S., MELO, W., AND THOMAS, W. Analytical and empirical evaluation of software reuse metrics. In *Proceedings of the 18th International Conference on Software Engineering* (USA, 1996), ICSE '96, IEEE Computer Society, p. 189–199.

[35] DIJKSTRA, K., PIETERSE, M. E., AND PRUYN, A. T. H. Individual differences in reactions towards color in simulated healthcare environments: The role of stimulus screening ability. *Journal of environmental Psychology 28*, 3 (2008), 268–277.

[36] DODD, N. G., AND GANSTER, D. C. The interactive effects of variety, autonomy, and feedback on attitudes and performance. *Journal of organizational behavior 17*, 4 (1996), 329–347.

[37] DRUCKER, P. F. The new productivity challenge. *Harvard business review 69*, 6 (1991), 69–69.

[38] DRUCKER, P. F. Knowledge-worker productivity: The biggest challenge. *California Management Review 41*, 2 (Jan 1999), 79–94.

[39] DUXBURY, L., HIGGINS, C., AND NEUFELD, D. Telework and the balance between work and family: is telework part of the problem or part of the solution? 218–255.

[40] EISENBERGER, R., JONES, J. R., STINGLHAMBER, F., SHANOCK, L., AND RANDALL, A. T. Flow experiences at work: for high need achievers alone? *Journal of Organizational Behavior 26*, 7 (Nov 2005), 755–775.

[41] FEINER, S., MACINTYRE, B., HAUPT, M., AND SOLOMON, E. Windows on the world: 2d windows for 3d augmented reality. In *Proceedings of the 6th Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA, 1993), UIST '93, ACM, pp. 145–155.

[42] FORSGREN, N. Octoverse spotlight: An analysis of developer productivity, work cadence, and collaboration in the early days of covid-19. *https://github.blog/2020-05-06-octoverse-spotlight-ananalysis-of-developer-productivity-work-cadence-and-collaboration-in-the-early-days-of-covid-19/* (2020).

[43] GOODHART, C. A. E. *Monetary Theory and Practice: The UK Experience*. Macmillan Press, 1984.

[44] GRAHAM, P. Maker's schedue, manager's schedule, 2009. `http://www.paulgraham.com/makersschedule.html`.

[45] GRANT, A., GINO, F., AND HOFMANN, D. Reversing the extraverted leadership advantage: The role of employee proactivity. *Academy of Management Journal 54* (06 2011), 528–550.

[46] GRAZIOTIN, D., WANG, X., AND ABRAHAMSSON, P. Do feelings matter? on the correlation of affects and the self-assessed productivity in software engineering. *Journal of Software: Evolution and Process 27* (08 2014).

[47] GROUT, C., ROGERS, W., APPERLEY, M., AND JONES, S. Reading text in an immersive head-mounted display: An investigation into displaying desktop interfaces in a 3d virtual environment. In *Proceedings of the 15th New Zealand Conference on Human-Computer Interaction* (New York, NY, USA, 2015), CHINZ 2015, ACM, pp. 9–16.

[48] GRUBERT, J., OFEK, E., PAHUD, M., AND KRISTENSSON, P. O. The office of the future: Virtual, portable, and global. *IEEE Computer Graphics & Applications 38*, 6 (January 2019), 125–133.

[49] GRUBERT, J., WITZANI, L., OFEK, E., PAHUD, M., KRANZ, M., AND KRISTENSSON, P. O. Text entry in immersive head-mounted display-based virtual reality using standard keyboards. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (2018), IEEE, pp. 159–166.

[50] GUILLOU, H., CHOW, K., FRITZ, T., AND MCGRENERE, J. Is your time well spent? reflecting on knowledge work more holistically. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2020), CHI '20, Association for Computing Machinery, p. 1–9.

[51] GUO, J., WENG, D., ZHANG, Z., JIANG, H., LIU, Y., WANG, Y., AND DUH, H. B.-L. Mixed reality office system based on maslow's hierarchy of needs: Towards the long-term immersion in virtual environments. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (Oct 2019), pp. 224–235.

[52] HACKMAN, R. J., AND OLDHAM, G. R. Motivation through the design of work: Test of a theory. *Organizational behavior and human performance 16*, 2 (1976), 250–279.

[53] HAIK, J., TESSONE, A., NOTA, A., MENDES, D., RAZ, L., GOLDAN, O., REGEV, E., WINKLER, E., MOR, E., ORENSTEIN, A., AND HOLLOMBE, I. The use of video capture virtual reality in burn rehabilitation: the possibilities. *Journal of Burn Care & Research 27*, 2 (2006), 195–197.

[54] HAYNES, B. P. An evaluation of office productivity measurement. *Journal of Corporate Real Estate 9*, 3 (2007), 144–155.

[55] HEKTNER, J., SCHMIDT, J., AND CSIKSZENTMIHALYI, M. *Experience Sampling Method: Measuring the Quality of Everyday Life*. 08 2006.

[56] HOLLINGER, V. Cybernetic deconstructions: Cyberpunk and postmodernism. *Mosaic 23*, 2 (1990), 29.

[57] HYMAN, J., BALDRY, C., SCHOLARIOS, D., AND BUNZEL, D. Work-life imbalance in call centres and software development. *British Journal of Industrial Relations 41* (02 2003), 215–239.

[58] IQBAL, S. T., AND BAILEY, B. P. Investigating the effectiveness of mental workload as a predictor of opportune moments for interruption. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems* (New York, NY, USA, 2005), CHI EA '05, ACM, pp. 1489–1492.

[59] IQBAL, S. T., AND BAILEY, B. P. Leveraging characteristics of task structure to predict the cost of interruption. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2006), CHI '06, Association for Computing Machinery, pp. 741–750.

[60] IQBAL, S. T., AND HORVITZ, E. Disruption and recovery of computing tasks: Field study, analysis, and directions. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (New York, NY, USA, 2007), CHI '07, Association for Computing Machinery, pp. 677–686.

[61] JACKSON, S., EKLUND, B., AND MARTIN, A. *The Flow Scales Manual*. California: Mind Garden Inc., 2010.

[62] JAKOBSEN, M. R., FERNANDEZ, R., CZERWINSKI, M., INKPEN, K., KULYK, O., AND ROBERTSON, G. Wipdash: Work item and people dashboard for software development teams. In *12th IFIP TC13 Conference in Human-Computer Interaction (INTERACT 2009)* (August 2009), Springer Verlag.

[63] JONES, C. Software metrics: good, bad and missing. *Computer 27*, 9 (1994), 98–100.

[64] KHAN, I., BRINKMAN, W., AND HIERONS, R. Do moods affect programmers' debug performance? *Cognition, Technology & Work 13* (2010), 245–258.

[65] KIM, H. K., PARK, J., CHOI, Y., AND CHOE, M. Virtual reality sickness questionnaire (vrsq): Motion sickness measurement index in a virtual reality environment. *Applied ergonomics 69* (2018), 66–73.

[66] KIM, Y.-H., CHOE, E. K., LEE, B., AND SEO, J. Understanding personal productivity: How knowledge workers define, evaluate, and reflect on their productivity. CHI '19, Association for Computing Machinery, p. 1–12.

[67] KNIERIM, P., SCHWIND, V., FEIT, A. M., NIEUWENHUIZEN, F., AND HENZE, N. Physical keyboards in virtual reality: Analysis of typing performance and effects of avatar hands. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2018), CHI '18, ACM, pp. 345:1–345:9.

[68] KO, A. J. *Why We Should Not Measure Productivity*. Apress, 2019, p. 21–26.

[69] KRANEBURG, A., FRANKE, S., METHLING, R., AND GRIEFAHN, B. Effect of color temperature on melatonin production for illumination of working environments. *Applied ergonomics 58* (2017), 446–453.

[70] LAKHANPAL, B. Understanding the factors influencing the performance of software development groups: An exploratory group-level analysis. *Information and Software Technology 35*, 8 (1993), 468–473.

[71] LATORELLA, K. A. Effects of modality on interrupted flight deck performance: Implications for data link. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (1998), vol. 42, SAGE Publications, pp. 87–91.

[72] LINN, M. C., AND PETERSEN, A. C. Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development 56*, 6 (1985), 1479–1498.

[73] LOK, B., FERDIG, R. E., RAIJ, A., JOHNSEN, K., DICKERSON, R., COUTTS, J., STEVENS, A., AND LIND, D. S. Applying virtual reality in medical communication education: current findings and potential teaching and learning benefits of immersive virtual patients. *Virtual Reality 10*, 3-4 (2006), 185–195.

[74] MARK, G., GONZALEZ, V. M., AND HARRIS, J. No task left behind? In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '05* (New York, New York, USA, 2005), ACM Press, p. 321.

[75] MARK, G., GUDITH, D., AND KLOCKE, U. The cost of interrupted work. In *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08* (2008), ACM Press, p. 107.

[76] MARK, G., IQBAL, S. T., CZERWINSKI, M., AND JOHNS, P. Bored mondays and focused afternoons: The rhythm of attention and online activity in the workplace. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2014), CHI '14, Association for Computing Machinery, p. 3025–3034.

[77] MARK, G., IQBAL, S. T., CZERWINSKI, M., JOHNS, P., AND SANO, A. *Neurotics Can't Focus: An <i>in Situ</i> Study of Online Multitasking in the Workplace*. Association for Computing Machinery, New York, NY, USA, 2016, p. 1739–1744.

[78] McGILL, M., BOLAND, D., MURRAY-SMITH, R., AND BREWSTER, S. A dose of reality: Overcoming usability challenges in vr head-mounted displays. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (New York, NY, USA, 2015), CHI '15, ACM, pp. 2143–2152.

[79] McINERNEY, C. R. Working in the virtual office: Providing information and knowledge to remote workers. *Library & Information Science Research 21*, 1 (1999), 69–89.

[80] MEYER, A., BARTON, L., MURPHY, G., ZIMMERMANN, T., AND FRITZ, T. The work life of developers: Activities, switches and perceived productivity. *IEEE Transactions on Software Engineering PP* (01 2017), 1–1.

[81] MEYER, A., FRITZ, T., MURPHY, G., AND ZIMMERMANN, T. Software developers' perceptions of productivity.

[82] MEYER, A. N., MURPHY, G. C., FRITZ, T., AND ZIMMERMANN, T. *Developers' Diverging Perceptions of Productivity*. Apress, 2019, p. 137–146.

[83] MOCKUS, A., FIELDING, R. T., AND HERBSLEB, J. D. Two case studies of open source software development: Apache and mozilla. *ACM Trans. Softw. Eng. Methodol. 11*, 3 (July 2002), 309–346.

[84] MURPHY-HILL, E., JASPAN, C., SADOWSKI, C., SHEPHERD, D. C., PHILLIPS, M., WINTER, C., DOLAN, A. K., SMITH, E. K., AND JORDE, M. A. What predicts software developers' productivity? *Transactions on Software Engineering* (2019).

[85] NEWPORT, C. *Deep work: Rules for focused success in a distracted world*. Hachette UK, 2016.

[86] NORMAN, G. Likert scales, levels of measurement and the "laws" of statistics. 625–632.

[87] PERRY, S., RUBINO, C., AND HUNTER, E. Stress in remote work: two studies testing the demand-control-person model. *European Journal of Work and Organizational Psychology 27* (06 2018), 1–17.

[88] PRITCHARD, R. D. Productivity measurement and improvement: Organizational case studies. Greenwood Publishing Group.

[89] PRIVETTE, G. Peak experience, peak performance, and flow: A comparative analysis of positive human experiences. *Journal of Personality and Social Psychology 45*, 6 (1983), 1361–1368.

[90] RALPH, P., BALTES, S., ADISAPUTRI, G., TORKAR, R., KOVALENKO, V., KALINOWSKI, M., NOVIELLI, N., YOO, S., DEVROEY, X., TAN, X., ZHOU, M., TURHAN, B., HODA, R., HATA, H., ROBLES, G., MILANI FARD, A., AND ALKADHI, R. Pandemic programming: How covid-19 affects software developers and how their organizations can help. *Empirical software engineering 25* (11 2020), 1–35.

[91] RALPH, P., AND TEMPERO, E. Construct validity in software engineering research and software metrics. In *Proceedings of the 22nd International Conference on Evaluation and Assessment in Software Engineering 2018* (New York, NY, USA, 2018), EASE'18, Association for Computing Machinery, p. 13–23.

[92] REBENITSCH, L., AND OWEN, C. Review on cybersickness in applications and visual displays. *Virtual Reality 20*, 2 (Jun 2016), 101–125.

[93] RUSSELL, H., O'CONNELL, P., AND MCGINNITY, F. The impact of flexible working arrangements on work-life conflict and work pressure in ireland. *Gender, Work & Organization 16* (05 2007).

[94] RUVIMOVA, A., KIM, J., FRITZ, T., HANCOCK, M., AND SHEPHERD, D. C. "transport me away": Fostering flow in open offices through virtual reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2020), CHI '20, Association for Computing Machinery, p. 1–14.

[95] SCHAFFER, O. Crafting fun user experiences: A method to facilitate flow. *Human Factors International* (2013).

[96] SCHNEIDER, D., OTTE, A., GESSLEIN, T., GAGEL, P., KUTH, B., DAMLAKHI, M. S., DIETZ, O., OFEK, E., PAHUD, M., KRISTENSSON, P. O., MÜLLER, J., AND GRUBERT, J. Reconviguration: Reconfiguring physical keyboards in virtual reality. *IEEE Transactions on Visualization and Computer Graphics 25*, 11 (October 2019), 3190–3201.

[97] SCHOLTES, I., MAVRODIEV, P., AND SCHWEITZER, F. From aristotle to ringelmann: a large-scale analysis of team productivity and coordination in open source software projects. *Empirical Software Engineering 21* (04 2016).

[98] SHARAR, S. R., MILLER, W., TEELEY, A., SOLTANI, M., HOFFMAN, H. G., JENSEN, M. P., AND PATTERSON, D. R. Applications of virtual reality for pain management in burn-injured patients. *Expert review of neurotherapeutics 8*, 11 (2008), 1667–1674.

[99] SNYDER, J., MATTHEWS, M., CHIEN, J., CHANG, P. F., SUN, E., ABDULLAH, S., AND GAY, G. Moodlight: Exploring personal and social implications of ambient display of biosensor data. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (2015), CSCW '15, pp. 143–153.

[100] SPEIER, C., VALACICH, J. S., AND VESSEY, I. The influence of task interruption on individual decision making: An information overload perspective. *Decision Sciences 30*, 2 (1999), 337–360.

[101] SPIVACK, A. J., AND RUBIN, B. A. Spaces to control creative output of the knowledge worker: a managerial paradox? Association for Computing Machinery, p. 312–318.

[102] STANNEY, K. M., HALE, K. S., NAHMENS, I., AND KENNEDY, R. S. What to expect from immersive virtual environment exposure: Influences of gender, body mass index, and past experience. *Human Factors 45*, 3 (2003), 504–520.

[103] SURALE, H. B., GUPTA, A., HANCOCK, M., AND VOGEL, D. Tabletinvr: Exploring the design space for using a multi-touch tablet in virtual reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2019), CHI '19, ACM, pp. 13:1–13:13.

[104] SWANN, C., PIGGOTT, D., CRUST, L., KEEGAN, R., AND HEMMINGS, B. Exploring the interactions underlying flow states: A connecting analysis of flow occurrence in european tour golfers. *Psychology of Sport and Exercise 16* (2015), 60–69.

[105] TARIS, T. W., AND SCHREURS, P. J. G. Well-being and organizational performance: An organizational-level test of the happy-productive worker hypothesis. *Work & Stress 23*, 2 (2009), 120–136.

[106] THOONDEE, K. D., AND LANE, C. Using Virtual Reality to Reduce Stress at Work. In *Computing Conference* (2017), no. July, pp. 492–499.

[107] TREUDE, C., FIGUEIRA FILHO, F., AND KULESZA, U. Summarizing and measuring development activity. In *Proceedings of the 2015 10th Joint Meeting on Foundations of Software Engineering* (New York, NY, USA, 2015), ESEC/FSE 2015, Association for Computing Machinery, p. 625–636.

[108] VALTCHANOV, D., BARTON, K. R., AND ELLARD, C. Restorative effects of virtual nature settings. *Cyberpsychology, Behavior, and Social Networking 13*, 5 (2010), 503–512.

[109] VISCHER, J. Will this open space work? *Harvard Business Review 77*, 3 (1999), 28–40.

[110] WAGNER, S., AND RUHE, M. A systematic review of productivity factors in software development. *ArXiv abs/1801.06475* (2018).

[111] WALSTON, C. E., AND FELIX, C. P. A method of programming measurement and estimation. *IBM Syst. J. 16*, 1 (Mar. 1977), 54–73.

[112] WANG, Q., XU, H., ZHANG, F., AND WANG, Z. Influence of color temperature on comfort and preference for led indoor lighting. *Optik 129* (2017), 21–29.

[113] WRIGHT, T. A., CROPANZANO, R., AND BONETT, D. G. The moderating role of employee positive well being on the relation between job satisfaction and job performance. *Journal of Occupational Health Psychology 12*, 2 (2007).

[114] XANTHOPOULOU, D., BAKKER, A. B., DEMEROUTI, E., AND SCHAUFELI, W. B. Work engagement and financial returns: A diary study on the role of job and personal resources. *Journal of Occupational and Organizational Psychology 82*, 1 (2009), 183–200.

[115] YAROSLAVSKI, D. Lightbot. *Retrieved on 25th May* (2014).

[116] ZHOU, M., AND MOCKUS, A. Developer fluency: Achieving true mastery in software projects. In *Proceedings of the Eighteenth ACM SIGSOFT International Symposium on Foundations of Software Engineering* (New York, NY, USA, 2010), FSE '10, Association for Computing Machinery, p. 137–146.

[117] ZIJLSTRA, F. R. H., ROE, R. A., LEONORA, A. B., AND KREDIET, I. Temporal factors in mental work: Effects of interrupted activities. *Journal of Occupational and Organizational Psychology 72*, 2 (jun 1999), 163–185.

[118] ZRAATI, P. Color consideration for waiting areas in hospitals. In *ICoRD'13*. Springer, 2013, pp. 1369–1379.