Location wise

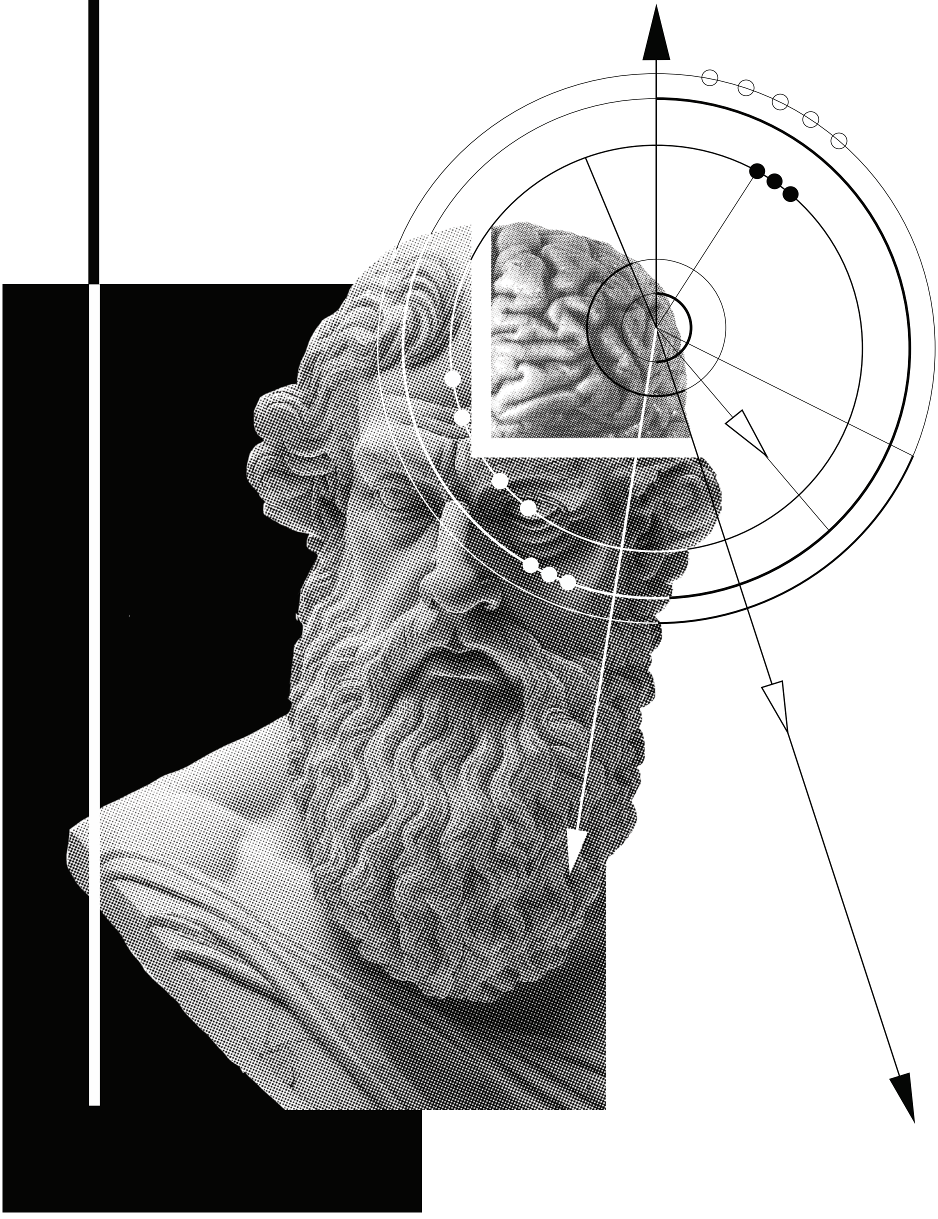
Document status and date:
Published: 01/01/2023

DOI:
10.26481/dis.20230310ms

Document Version:
Publisher's PDF, also known as Version of record

Location Wise

Behavioral Location Decisions



Martijn Stroom

Location Wise

Behavioral Location Decisions

Martijn Stroom

ISBN: 978-94-6469-264-8

Cover by Martijn Stroom (concept), Joëlle Martin (design), and Midjourney AI (sculpture).
Printed by ProefschriftMaken.nl

Maastricht University

Location Wise

Behavioral Location Decisions

Dissertation

To obtain the degree of Doctor at the Maastricht University, on the authority of the Rector Magnificus, Prof. dr. Pamela Habibovic in accordance with the decision of the board of Deans, to be defended in public on Friday 10th of March 2023, at 10:00 hours.

by

Martijn Stroom

Promotors

Prof. dr. Piet Eichholtz

Prof. dr. Nils Kok

Assessment Committee

Prof. dr. Rachel Pownall (Chair)

Prof. dr. Caroline Goukens

Prof. dr. Jörg Gross (University of Zurich)

Dr. Peiran Jiao

Prof. dr. Michael I. Norton (Harvard Business School)

"'What gets us into trouble is not what we don't know. It's what we know for sure that just ain't so."

Mark Twain

Acknowledgements

My experience from all the wonderful dissertations before me led me to two certainties. First, 99% of the people I hand this book to, will not read beyond this chapter (and afterwards, I'm sure my parents wished they didn't either). Second, it is inevitable to forget names and stories I wish I included the moment I receive the printed version. I hope you will understand that I have appreciated all moments, especially those I don't remember. A couple of months ago, Piet and Nils gifted me a book called 'brevity', hoping (and urging) I would write less lengthy and more concisely.[1] Needless to say, and evident from this chapter, I did not read it yet. I hope they will forgive me.

More than anybody else, it's safe to say I would not be here without my two supervisors. Piet, if it wasn't for you, I would have never started this PhD position, let alone make it through the first year. Whenever I get asked how I got into the real estate team, I enjoy telling them how I did not apply for this position, I rejected it when you offered it to me, and how I tried to quit after six months when circumstances got the better of me.[2] In the last two situations, you simply did not let me. With anybody else but you, that would have meant the end of my academic carrier. For this, I will always be thankful. In the best way possible, conventional rules and regulations don't seem to apply when you are around. What one imagines, you enable. It will come as no surprise that the origin of the (now publicly acknowledged) expression "getting Pieted" lies partially with me. In my view, this expression encapsulates you completely: you don't only seize opportunities, but you create them, even for people who never knew they aspired them in the first place.

Nils, if it wasn't for you, I would have never finished. After your return from the US, our contact increased, and one could argue that the correlation with my output might be partially causal. Your swift alternation between aspiration and practicality makes you a motivational leader. I think it is impossible to match your energy, and keeping up (much like 'having a casual walk' with you) can be challenging at times, but I enjoy trying. You offer room to grow where needed, create possibilities on-demand with unmatched enthusiasm, and introduce shock-therapy style responsibilities. Whenever I evaluate a new idea I have, I often hear your voice in my head ("Ah, you know… I like it").[3] Whereas the match with finance was sometimes

---

[1] I am still awaiting a book on how to say no. Suspiciously, this book seems to have gotten lost in the mail.
[2] Of course, I include in the story that I applied for a pension decision aid, and my whole application letter consisted of me explaining how important pensions were and what I would like to do on the project. Evidently, you never read the application letter.
[3] Granted, I hear it more often in my head than in real-life.

challenging, I have never doubted the match of team. You both possess bravado, confidence, and can-do mindsets, exactly what I needed to develop when I walked in.[4] It has always been my belief that, even if my academic growth would have been marginally suboptimal from an expertise perspective, the personal growth alone would be worth triple the investment.

Finally, my foster supervisor, Martin. Funnily enough, also you played an important role in starting this position, as your advice and recommendations have motivated me to apply to finance and accept a meeting with Piet. I'm sure you underestimated how often I would come by to ask trivial things like how to pay my participants. You are notoriously bad at responding to emails or making meetings on time, but when I walk by, you clear your table, take out one of your weird chairs, and free up all the time I need.[5] I remember sitting on the kitchen floor of a party on a random Saturday night at 3 am sending you an email trying to solve a problem with our paper, to which you responded immediately, telling me at least one of us should not be working, and it better be me. I was proud to have co-authored my first publication with you, and more great things are ahead with our current working paper.

Many people have been pivotal in the last years. Nevertheless, four special selections support my defense as paranymphs. Lidwien, many years ago I knew that if I would ever get to defend you would be my paranymph. Although you seem to brag about how scared I was of you in the beginning (not a compliment, I would say), once we started talking, we almost immediately became inseparable. With similar sentiments on academia and academics, we have achieved many great things together. None of them were academic. In a weird way, you simultaneously made me want to quit and persevere at the same time. Kasper, I probably leaned on you most out of all my nonacademic friends during the PhD time. You kept me sane, level-headed, and often under influence. You made sure that I remembered that there was an actual world outside the university, and the skills I acquired gave me absolutely zero authority anywhere. As the PhD transition coincided with personal transitions, I often think I could have lost myself in the PhD without you. All the adventures we had, including Sunday morning K&D evaluation calls as well as personally importing COVID-19 into the Netherlands, made my life so much better. Hugo, the soft, kind, and chivalrous attitude hides your mischievous nature. If only people knew how good you are at imitating everybody. A great host, engaged community member, and excellent mechanic. You are one of the most liked people in our department and

---

[4] Some of you might have difficulty not laughing seeing me infer 'lacking confidence'. Please take a minute to catch your breath. I would merely argue that professional confidence is different from Platielstraat confidence. The distinctions I found are that the former seems totally independent of increasing wrinkles, whereas the latter is completely useless in academia.

[5] Which, I am sure you know, is exactly the reason you are often late and don't have time to respond to emails.

for good reason. Finally, Linde. My appreciation for you is based on two kinds of respect. First, I think you are killing it in your PhD. Seldom does somebody convince me as well as you do, especially on behavioral issues. Your ability to manage other team members also deserves compliments. But more importantly, I respect your ability to bully me. The moment I realized that during dinner I was still thinking about something you said over lunch, was the moment I realized you were cool.

Alma mater literally translates to nourishing mother and normally refers to the school you graduate from. For me, however, almae matres (plural) refer to the actual women at this school that have made this PhD so much more enjoyable. First, the FIN secretariat has provided me with the knowledge, funding, support, and gossip I needed. Without a doubt the stable factor throughout these years: Francien, a.k.a. Frenske. Thank you for all the support, suggestions, and relationship advice. I evidently ignored all of them but appreciated them nonetheless. Carina, we started with parent-teacher meetings and ended with Iphone smuggling. What a ride. Texting with you in Mestreechs makes my autocorrect mad, but me happy. Monique, I'm sorry for the many requests, but at least I enjoyed that you came to visit me in Lisbon. And of course, Cecile and Els cannot be forgotten. Second, a big thanks to Nicole and Pascalle from the Marketing secretariat. You have always made sure I was well-fed with whatever food was left. And you never asked anything is return, the true testament that altruism does exist. Third, the ladies from the front office, Ellen, Desiree, and Yvonne, for often covering my mistakes with the cloak of charity. During the pandemic, you were always there to have a little chat at the end of the day. Finally, but certainly not least, my foster-mother-in-law, Lyan. I truly feel like your son-in-law. Our hikes with Rommel, dinners with AS and Max, Lisbon planning, and many talks about life and work always relaxed me during stressful periods. You made Epen feel like my summer house. The support you offered is limitless and exceptional, and I consider myself very fortunate that you and Max are such loving and caring people. Thank you for everything.

I started my PhD in B1.08. This room has been the cradle of many great names. In my early days, I needed some time to adapt, but Juan and Matteo offered all the support I needed. The unity of this office is something I look back at with pride. For example, whenever one of us presented somewhere, at least one of us would be there. I still laugh remembering Matteo, sitting in the front row, on a Saturday afternoon, at Randwyck, listening to me talk about psychology during a psychology master's open day. These things might seem trivial but are exemplary of these two men. Juan, you are a gentle character with vast amount of knowledge and open-door policy. You are starting to reap what you sowed and deserve, and I am happy to see you return. Matteo, we had many adventures which at least one of us remembers, and I am looking forward to our next! Even in your absence, your picture on my office wall has supported me through tough times. Katrin, I am thankful for all the talks we have, the advice you gave me, and the enthusiasm you continue to cast upon me. You started in my office and I ended in yours, which led me to Marco. This Italian, Swiss, Romanian, hardworking, charismatic gentleman will leave soon, but not before bestowing his empathy and knowledge upon me, like the gentle hand on my shoulders whenever he enters. Finally, I left the office during the era of prodent-smile Alex. I think we did pretty well with our side-quest on floodings, and it has been nothing but pleasure working with you.

The real estate team is a powerhouse of talent. Mike, I am generally squeamish to live together with somebody, but not with you (it helped you were never there). You are a hard worker yet still aspire to live your best life. The moment the pitch of your voice goes up, your credibility goes down. And I've always secretly admired your bravery, you endeavor adventures like you eat breakfast. Nico, our time together was short, but sweet. I wish every meeting would end with chatting with you. I am very much looking forward overstaying my welcome in the future. Stefan, our trip to Hawaii felt like a dream, almost unreal. I am happy that you joined as an experimentalist, as we share many opinions, experiences, and German jokes. I admire how you formalized what you want along the way, and I know you will get there. Dongxiao, as a relatively recent addition to the team, you have quickly won over the hearts of us all. You are very caring, engaged, focused, and easy to hang around with. I have no doubt your next step will be huge. AleX, your big smile during tough times shows your perseverance. You will get there. Minyi, you initially appeared relatively shy (and seemed to be cold, given the jacket), but looks can be deceiving. I am looking forward to your next steps.

I've often felt that a PhD is a bitter sweet period. Slowly but steadily, you say goodbye to the PhD's you started with until there is no one left. I need to include some special mentions from my old cohort. Inka, as a proud member of hyggelig app, I am happy we are still in contact. You are exceptionally strong, and finally close enough to meet in person again. Addy, I have hugely appreciated our walks through the park and coffees during our weekend office sessions in my first year. You are kind by heart and (relatively) wise, and I cannot be happier you ended up exactly where you wanted to be.[6] Tobias, unbeknownst to you, our talk at your defense was the first moment I started to believe I might actually finish. From that moment onwards, your thesis has been my bible. Ming, we often joked that I would finish before you. Neither of us

[6] Not in Kenya

believed it. I enjoyed making you swear in public, and are happy you found your spot. Nora, one of the brightest and modest PhD's I've met, you always impressed me. Our adventures group excursion in Winterberg and LA were so much fun. Marina, you are to date the only person that every got me kicked out of a party, and I am sure this makes you very proud. Technically new, but I still count you as old: Marten and Colin. Just before the pandemic we had the pleasure to explore Mexico, and I cannot think of better companions: Colin as the whip, and Marten as the agreeable tiebreaker.[7] Colinski, a.k.a hightower, you would have excelled in academia without ever working on the weekend. You are doing great in Holland, but I often miss you in the office. The quality of my jokes is hardly ever questioned, and imitating you is less fun when you don't see it. Marten, a.k.a. mini-me, although you haven't officially left yet, I also often miss you in the office. Our matching clothing style is now completely out of sync. You are so successful in creating an awesome work-life balance with your wife Janine that you are a true example to us all, with an exciting new chapter on the horizon.

In the beginning, I often felt the need to explain to new PhD's that it's normally (or previously) not as fun in an academic department, to avoid disappointments. I was wrong, as the department seems to get more social every year. Each of you are so social that it is almost unacademic. Some honorable mentions included the party committee of Jonas, Flavio, and Samara. Clearly, you have introduced a next level of social comradery. Robinski, I cannot wait until I meet you randomly in 10 years, successful, so I can tell you that I told you so. Your bat-like audio location mechanism makes me laugh every day.[8] Bin, you are the only person I know that tells other people about my papers besides me, thank you! The downside of taking so long to finish, is that the list of colleagues to thank is almost endless and arguably longer than my actual dissertation. Thus, to all old and new FIN PhD's: thank you.

Over the years, I've also grown increasingly attached to our senior department. Thomas, your attitude makes me smile every day. I initially applied with you, and I am sure I would have enjoyed it as much as I enjoyed our behavioral finance course and our carnival trips. Maybe we still get to do a cool project at some point, I know where to find you. Paulo, my favorite behavioral scientist. Our discussions during lunch have been a treat, as you are as convinced of your stubborn opinion as I am. You are one of the easiest people to walk into the office, although you often lose me after 2 minutes[9]. That you nonetheless continue to try, is

[7] In case you wondered, Marten snores like a full-grown man. Janine apparently says he doesn't which only begs the question what is wrong with her. It's the kind of snoring that supersedes mere audio and is at the seismic waves level. Also, Colin hypothesized that monkeys are put on islands for tourists. To this day, this has been the only time he had an original hypothesis not previously published.
[8] For those unfamiliar with Robin's soundboard, please visit her office.
[9] I think we could blame my workshop in finance teacher for that.

admirable. Dennis, after 3 years of sitting next to you during the seminars, you finally learned my name. The pandemic has made us spend almost all lunches together, and it's safe to say that they became increasingly more enjoyable. Thank you for all the effort and comments you provided on my work. Jaap, for most of my time here, you were the fierce leader of the gang. You are able to alternate jokes with valuable advice, and have provide me with both. Rachel, your kind words and little chats are always enjoyable. Sjoke and Sanne, our Mestreechter bond is undeniable. Kimon, we went from pizzadelivery guys to SBE colleagues: couldn't be better. Jeroen, your sustainable finance course laid the groundwork for being able to survive in the finance department.[10] Paul, your enthusiasm and our many brainstorm sessions on sexy projects emphasized the joy there is to be had in research. Peiran, your knowledge and performance as formal behavioral economist are inspiring. Pumske, through it all, you always had a big smile and empty mug of coffee ready to be filled. Stefan, Stephanie, Peter, Dirk, and Joost: thank you all.

Also outside the finance department are some noticeable mentions. Ingrid, you are always ready to provide me with both academic and relationship advice. I have high hopes for both. You and Burak are always supportive and show endless belief in my ability, for which I cannot thank you enough. Roselinde, working with you has been, and still is, a true pleasure. Your efforts have enabled me to evolve into my role as project lead. Melline and Anna, you have made it secretly impossible for me to keep making fun of ROA. I still try though, and I'm sure Anna especially needs to endure that for still some time. Eric, our coffees at Koffie, data discussions, and surfing holiday taste like more. The PhD committee members as well as the Science Slam organization have additionally been extracurricular activities that felt like a hobby. Thank you, Hendrik, Kim, Steffi, and all others from that period. Suzanne, much like Martin, you have been instrumental in getting to this position. Without your talks and advice, I would have never gotten to this moment. Thank you.

Although I increasingly neglected you, I owe most to my friends. Hands down the most effortless relationship I ever had has been with Donny. With a name like you are from Wittevrouwenveld, you were easily the most stable factor in my daily life. We laugh more than we work-out, and even in my most frustrated periods, I walk out of the gym with a big smile on my face. If I would have spent more days in the gym without you, I would have had a sixpack and Kel would have her boyfriend home more often. I think we are currently both happier. Honestly, we are a podcast waiting to happen. Sammy, one of my oldest friends, even if we don't

[10] I've watched your narrated course video's so many times, that at some point you were narrating my dreams.

see each other for some time, you are always just one call away. The way you burden yourself to take care of those around you is something else. Sliks, my little sister, I am thankful for all the times you forced our talks and walks, our Christmas dinners, and generally took care of me. You know what I think, and I am sure of it. Rickert, you are the only one that owes me more than I owe you (specifically a working car and three winter holidays on skies). I will haunt you until your debt is paid. In the meantime, I hope we can keep up our ski and golf traditions. Jenna, your art sharing combined with my psychological analysis often throws us down a rabbit hole, which I thoroughly enjoy. Shannen, thank you for literally making me a bit less criminal. I cannot forget about Audrey, Gilbert, Marije, Steffi, Stephi, Emilia, Rosijntje, Caro and Lukas. Finally, the commando's[11]: Ray, Nick, Willem, and d'n Dave and the other boefjes. From complex evenings to wine tastings and woodshop visits, from painting doors at 3am to kids' birthdays and other key moments. From raising the roof at parties to staring at them, our curfew sleepovers, and endless deep and sophisticated discussions are the stories I will tell my children about (or yours). I cannot stress enough how our adventures have been the spice of my life. Bijzijn is meemaken.

Finally, my family. Pap en mam, zonder jullie was ik niet wie ik nu ben[12]. Er is veel veranderd in de afgelopen jaren, maar de wetenschap dat er altijd een plek is om terug te vallen maakt het mogelijk om avonturen aan te gaan. Mam, jij bent opportunistisch, sociaal, zacht, maar ook krachtig en laat je niet zomaar iets zeggen. Sommige kwaliteiten bezit ik, andere aspireer ik nog. Vanaf kleins af aan heb jij geprobeerd mij uit mijn comfort zone te duwen, en dat moet met zo'n professioneel moederskindje als ik ben niet makkelijk zijn geweest. Pap, strijdvaardig, maar vooral enorm begaan en zorgzaam. Als ik je aan anderen moet omschrijven, stel ik altijd dat jij met vliegangst nooit gaat vliegen, maar als er ooit iets gebeurt met mij aan de andere kant van de wereld, loop jij nog voor mam het vliegtuig in. Het kritische, het vurige discussiëren, en (over)denken (inclusief het hardop tegen jezelf praten) heb ik onomstotelijk van jou. Kim, mijn altijd trotse zus die na decennia inmiddels mijn rare eigenschappen heeft geaccepteerd. Na zwaardere tijden te hebben overwonnen ben jij een uitzonderlijk goede moeder. Ik ben wellicht nog trotser op jou dan andersom.

Thank you for playing such a crucial part in this period of my life.

Martijn Stroom
February 2023

[11] For those of you ready to anonymously report a secret militia, this (whatsapp group) name stems from the Dutch television series 'Jiskefet': "I've see a lot during my time with the commando's" *("Ik heb veel meegemaakt bij de commando's,..")*

[12] Both nature and nurture

Contents

List of Figures

List of Tables

Chapter 1

Introduction

"I wish there was a way to know you're in the good old days before you've actually left them." – Andy Bernard.

This timeless quote resonates with many and typically refers to the nostalgic value of a past period in your life. The actual memories might differ for each individual, but the emotional nostalgia experienced is highly identifiable for most. Playing carelessly on the soccer field as a child, enjoying a great night out as a student, or owning your first house, are all prime examples. At one point, often unbeknownst to yourself, such a period regresses and quietly slips into memory. At the start of Q2 2020, the world collectively found itself nostalgically longing back to better days.

The global COVID-19 pandemic had an immense impact on our lives and the way we were able to live it. Location became a key focus to combat the pandemic (Chinazzi et al., 2020; H. Fang et al., 2020). In 2019, location decisions mostly concerned a (retrospectively) luxury level – what country will I travel to, in which city will I work and live, and in which area should I buy a house. In 2020, those decisions seemed a thing of the past. After discovering that COVID-19 can be transmitted by getting within 6 feet of someone who is infectious, limiting social contact was of the highest priority[13]. Hence, in 2020, decisions about location pertained to daily movement – will I go outside, will I work from home, and will I meet up with others?

The degree to which movement had swiftly become strictly regulated and prescribed was unprecedented. Society was tested on its ability to take an economic and social hit. Resilience, flexibility, and trust in government were pushed to limits, as restrictions were often ill-understood (Devine et al., 2021; Mækelæ et al., 2020). The societal consequences of individual actions were not easily overseen and they consistently conflicted with personal wants and needs. Luckily, whilst finalizing this thesis, better days are on the horizon. Society as a whole has been able to knuckle through hardship and is anxiously ready for better days to come.

[13] Although 6 feet might have not been enough (Bazant & Bush, 2021).

Why do restrictions in movement seem to be such a hard pill to swallow? Clearly, our generation has never witnessed something comparable to this global equivalent of a ball and chain. Without a doubt, the mere effect of restricting movement in any context has great implications for quality of life and well-being (Inglehart et al., 2008). Experienced freedom is often defined as the freedom of choice, including the choice to go wherever you please (Barlas & Obhi, 2013; Schwartz, 2000). Growing up, freedom entails the development of self-determination to move around alone, deciding to go to (undisclosed) locations connecting to new social networks, often unbeknownst to your caretakers. The impact of the choices gradually evolves to more trivial life scenarios with age. For instance, you have the freedom to choose and change jobs as an adult. Flexibility on where to work on a daily basis might have a great influence on this decision. In contrast, the most threatening tool a civilized society has to punish people is to take away their freedom of location choice by drastically restricting and dictating movement.[14] Clearly, and not unlike real estate most famous quote suggests, location determines the value of life like it determines the value of your home.

What can we learn from our experiences in these times of movement deprivation? How do we evaluate ourselves coping with restrictions of location? This dissertation sets out to explore human behavior and thought processes with regards to (forced) changes in location decisions. I use the pandemic to give context to most of my research and my participants, but never as a unique prerequisite for my results. The pandemic aids me in unraveling underlying existing processes and preferences by forcing decisions, forcing behavioral change, or introducing uncertainty.[15] Together, we will explore how people deal with completely novel yet forceful policies, working from home, and (re)connecting with risky social networks. Although these situations are varied and differ fundamentally, they have the decisional context in common: location, location, location![16] Thus, let us consider three daily location decisions during the pandemic: will I go outside, will I work from home, and will I meet up with others?

[14] This is not the most effective psychological weapon society has. Baumeister's work dictates that social exclusion has the same mental effect as physical torture. Note that in prison only the size of potential contacts is limited, hence not strict social exclusion per se. Arguably, the group cohesion in prison can be enormous (for instance, think of the Stanford Prisoners Experiment; (Haney et al., 1973). However, the complexity of complete social isolation leads us to the second best.

[15] Like the societal resilience, I make the best out of the worst.

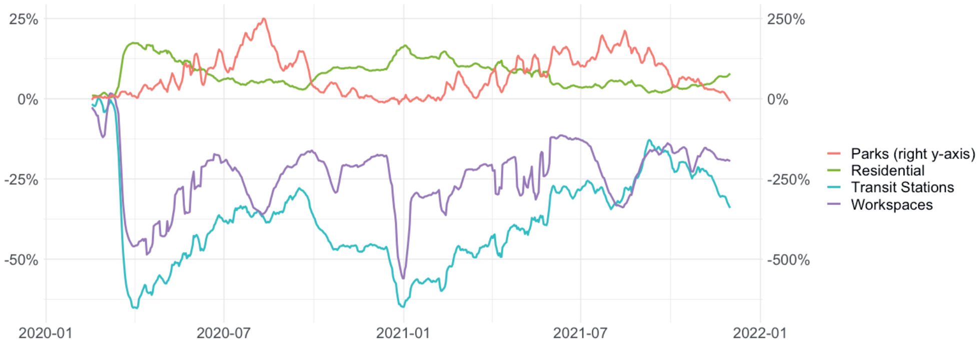
[16] This refers to the phrase "Only three things matter in real estate: location, location, location!". The quote, often ascribed to Harold Samuel in 1944, suggests that the value of a real estate asset is mostly determined by its location. Although, in the current and future climate crisis, we should maybe adjust it to "future location, future location, future location".

1.1 Stay or Go

The decision to go outside was the first immediate hurdle in our lives to overcome following the first global lockdown. Many policies clearly dictated and enforced a national quarantine at the start of the pandemic. These swiftly evolved in limitations in group sizes and curfews. The Netherlands reverted to one of the least forceful policies in the EU, and the decision to go out or not was effectively left with the population (RIVM, 2020). Accompanied by a frequently updated recommendation (stay home as much as you can or avoid busy places), each individual could (mostly) decide for themselves how to interpret the recommendation.

Although office, government, and retail buildings were mostly closed, the first COVID summer was riddled with small-scaled decisions to leave the house in search of fresh air and distraction. In the summer of 2020, 'stay at home' officially changed into an 'avoid busy places' recommendation. Figure 1.1 shows how Dutch citizens changed their location decision during the pandemic. People spent over 50% less time at their office and in summer increased park visits by almost 250% compared to the year before! Although this signaled positive news as people were following the advice to work from home whilst still getting fresh air, a question arose: did we still successfully avoid busy places? Unfortunately not. This was painfully illustrated by a picture of a Dutch beach that made worldwide headlines in the summer of 2020 ("Beach chaos threatens Europe as temperatures rise"; BBC News, 2020). Was this violation intended, or did it result from the complex strategic decision people had to make?

Figure 1.1 Number of visitors in the Netherlands relative to the period before the pandemic?



Source: Google COVID-19 Community Mobility Trends – OurWorldInData.org/coronavirus

Chapter 2 examines the effect of context on the decision to visit a hypothetical recreational hotspot under the policy recommendation to "avoid crowded places." Based on the observed violations of the recommendation in the Netherlands, I hypothesize that the absence of relevant up-to-date information about crowdedness will force individuals to make a decision based on unknown information, making it susceptible to biased reasoning. In essence, whether you are able to successfully follow the recommendation to avoid busy places depends on your belief in other peoples' choices. A location not visited by others can safely be visited, but if others reason similarly, should be avoided.[17] Thus, strategic decision-making will motivate most people to go out when they expect others to stay home, inevitably leading to escalation and subsequent failure to avoid crowdedness.

In the experiment, I randomly present four levels of context on the crowdedness "on the streets" (no context, low, medium, and high crowdedness context) to see if and how people use this context to inform their decision. Although there is a strong inclination to avoid crowded places during the COVID-19 pandemic (81%), I find two context-driven exceptions: when people expect to avoid crowded spots (in the "low" context; strategical decision-making) and when people expect others to go (social influence). The effect of education supports the underlying theory, as a higher educational background is more important in the rational or strategic (low context) decision than in the escalation (high crowdedness context). Hence, in the low context situation highly educated people act strategically, yet in the high context, social norms lead to escalation. The freedom provided by the ambiguous public policy is implicitly asking more from the population than it initially seems. "Use your common sense" is often the accompanying advice, but I show that more and better information concerning the context is essential to enable us to make optimal decisions for ourselves, and for society.

1.2 Work from Home (WFH)

After the initial shock response, the adaption of work and society to the pandemic quickly evolved. Figure 1.1 shows that office movements dropped immediately by 50% and

[17] This reasoning continues circularly: if others similarly conclude a hotspot should be avoided, you can go again, and onwards. The reason people don't just simply stay home to avoid any trouble is described by game theory. Simply put, common knowledge of rationality dictates that two rational people who need to coordinate will understand that the other person is as aware and as knowledgeable about both themselves and the others. In our example, the chain of reasoning (known as orders of beliefs) would continue indefinitely. Findings in the field show however, unsurprisingly, that we are irrational beings. In general, we do formally acknowledge that others are reasoning similarly to us (first order belief) and that they are aware we know this (second order belief). But we tend to stop there: by means of overestimation, we almost always think we are one step smarter than our opponents. It's this overconfidence that fuels our strategic thinking.

remained at minus 25% compared to pre-pandemic levels ever since. Non-essential offices' had severely limited opening hours or were closed altogether. Beyond all general uncertainty about the future, it was clear that this was not going to change in the short term. Working from home was to be the new default. Contrary to many other changes, however, working from home was not new. Albeit on a smaller scale, at least partially working from home had already risen in popularity in the decades before the pandemic.

The previous experience with working from home (WFH) predominately showed increasing job satisfaction (Bloom et al., 2015). And there is much to like. The trust employers have in their workers to let them work from home improves the self-management and self-motivation of those workers (Standen et al., 1999). Effectively, WFH job satisfaction grows through improving workers' psychological needs for autonomy, competence, and relatedness (Brunelle & Fortin, 2021)[18]. It further facilitates a more hands-on approach to family matters such as caregiving to parents of children, thus increasing the sense of general (interpersonal) belongingness (Vittersø et al., 2003). The lack of a commute enables one to spend less time on the road and more time with family or friends whilst at the same time decreasing the environmental footprint[19]. Not having to commute also increases geographical freedom. Choosing where to live was historically constrained to the distance to one's office, forcing many into expensive urban areas. Without the (daily) commute, the suitable residential radius expands, likely decreasing costs and increasing the amount of green one can see from the bedroom window (Gupta et al., 2022; Kahn, 2009).

WFH also came with some disadvantages. First, not all jobs are suitable for WFH. Generally, white-collar jobs are more suitable than blue-collar jobs. Second, communication and coordination with coworkers suffer as all encounters need to be scheduled specifically (Gajendran & Harrison, 2007). Bad communication in turn may lead to bad performance. Third, WFH makes it increasingly difficult to separate work from home (B. Wang et al., 2021). One reason for this is that workers at home feel the need to show that they deserve to work there (unsupervised), and thus overcompensate. Effectively, they feel they owe it to their employer to work extra hard (Felstead & Henseke, 2017). Enduring that extra effort for a long period of time could lead to burnout symptoms (Golden, 2006).

But the greatest caveat mentioned in WFH research is not practical or theoretical, but methodological. The self-selection into WFH has been the most substantial critique of positive

[18] That is, of course, when you are able to detach yourself from the work (Park et al., 2011).
[19] Humble brag benefits

WFH evaluation. Maybe employers who allowed WFH were in general more likely to elicit great job satisfaction and productivity scores. And workers evaluating WFH almost exclusively opted for WFH.[20] Thus, evaluations on satisfaction and productivity are likely to be high, as otherwise, those specific workers would not work from home (in behavioral science also known as the survivor bias[21]). In contrast, COVID-19 forced all workers to work from home, whatever their and their employers' preferences, and opened the possibility for objective evaluation.

1.3 The Productivity Questionnaire

In order to accurately assess the productivity of homeworkers, I first construct and validate a work satisfaction scale based on all known factors that could influence WFH differently. It is clear that the move from the office to the home environment was novel, forced, and abrupt. At a time when employees' physical and mental states were tossed around by the pandemic, employers were unable to monitor general satisfaction, well-being, and productivity. This increased the demand for periodical check-ups in the experienced productivity (Farooq & Sultana, 2021; Galanti et al., 2021; Zito et al., 2021). Luckily, the Health Work Questionnaire (Halpern et al., 2001; Shikiar et al., 2001, 2004) covers both productivity and its relation to health. Although cited by many well-known papers, my attempt to validate this questionnaire did not go as well as I hoped (Escorpizo et al., 2007; Healy et al., 2016; Hovinga et al., 2008; Lofland et al., 2004; Thorp et al., 2014). The initial validation by Shikiar further suffers from limited sample size and has been critiqued for its lacking connection to absenteeism (i.e. the unplanned absence from work due to illness or lacking concentration; Lohaus & Habermann, 2019; Mattke et al., 2007; Ospina et al., 2015). For a metric so pivotal in performance research, this productivity questionnaire aims to be better-identified than its current alternatives, yet still easy-to-apply.

In the first part of chapter 3, I conduct a novel principal component analysis, which results in a new questionnaire I label the work productivity and stress questionnaire (WPSQ). This questionnaire identifies five key subfactors: productivity, productivity as seen by others,

[20]Some papers rely on a semi-random allocation of WFH. I argue that this is still relatively biased, as employees are unlikely to send workers home that strictly did not want to. As such, it was more likely a random selection out of a pool at least open to WFH. Second, even a short-term WFH pilot could be more positively scored based on novelty and the expectation that, at a later stage, WFH would become merely optional.

[21] The survivor bias stems from the WOII evaluation of fighter planes returning from missions over German-occupied Europe. Initially, the planes were checked on damage, and the most hit areas were fortified so that they could take more hits and increase the likelihood of safe return for a next trip. However, that approach was crucially flawed as the planes returning survived because they were not hit at the other areas. Not until the barely hit areas on the plane were fortified, did the return rate of fighter planes increase. For WFH, focusing on the survivors of WFH programs is unwillingly overestimating the value in the general population by ignoring the preferred office workers.

stress and irritability, peer relations, and nonwork satisfaction. These factors outperform the six-factor solution proposed by the HWQ on factor fit and internal consistency. Since I collect data at two moments in time, I can also show internal consistency over time. In addition to the questionnaire, I measure and discuss a single-item scale of productivity and single-item approximations of absenteeism. Overall, chapter 3 shows that the newly formed WPSQ factors outperform the HWQ on a large sample based on internal consistency and reliability on two measurement periods and are thus fit to use for further research in chapter 4. Additionally, I conclude that the single-item scale alternatives could substitute some of the WPSQ factors but only with caution and only when brevity demands it.

1.4 Working from Home Productivity

Chapter 4 sets out to investigate which home office characteristics underlie WFH satisfaction, productivity, and willingness to continue after the restrictions have been lifted. My first findings show that people report higher productivity, satisfaction, and happiness with their work at home compared to their office work. Additionally, during WFH, burnout propensity decreased, coworkers' relationships increased in quality, and non-work satisfaction improved. But what causes this overall increase in satisfaction?

Let us turn our focus to the obvious yet least discussed difference between home and the office: the actual physical environment. As we move to our own home, we have the luxury to design it as we please. And this should not be taken lightly. Many innovative employers spend millions on designing a work environment at the office that makes workers productive and healthy, and makes them want to be at the office by making it feel like home (Al Horr et al., 2016; J. G. Allen et al., 2016; MacNaughton et al., 2017). And what feels more like home than home? On the other hand, the investment made by employers is not easily matched by employees – think about (standing) desks, advanced computer screens and, importantly, high quality internet connections. Do workers actively optimize their WFH environment in a way that increases satisfaction and supports productivity?

I show that indoor environmental quality (IEQ) factors like temperature and air quality are preferred at home, whereas physical attributes like desk and screen are preferred at the office. This is not surprising, since optimizing ergonomics at home can be challenging and often costly (Davis et al., 2020) whilst being in control of opening windows and the thermostat at their own preference is a skill that many (not all) master quickly (Chang & Kajackaite, 2019). But more importantly, I show that satisfaction on both indoor environmental quality as well as the physical environment is correlated with job satisfaction and productivity. Since indoor environmental quality scores higher at home, it must have a big impact on WFH success.

In the final part of chapter 4, I set out to relate actual actions of indoor environmental improvement by workers to indoor environmental quality satisfaction. Ventilation is key in influencing indoor environmental quality (J. G. Allen et al., 2016). Although unable to measure the actual indoor environmental quality, I require people to state how often they ventilated during work hours. By means of mediation analysis, I show that more ventilation does improve not only indoor environmental quality satisfaction but also the rationally unrelated physical attribute satisfaction. In essence, better-ventilated rooms increase overall satisfaction, even beyond the awareness of the workers. By improving overall satisfaction in the complete physical environment, more ventilation improves job satisfaction, job productivity, and willingness to continue to work from home. 'Hyperventilation' is for offices, in contrast to humans, hugely beneficial for general satisfaction, WFH success, and productivity.

1.5 The value of self-report in recollection of WFH productivity

The fact that factors such as ventilation influences satisfaction on completely unrelated issues, such as desk and chair satisfaction, opens a new discussion. How valuable would the results from chapter 4 have been if I would have focused just on self-reported satisfaction? Without ventilation, a likely conclusion would have included focusing on increasing satisfaction of factors that had significant room for improvement (e.g. the screen, desk, and Wi-Fi). Although psychology repeatedly warns about the validity of self-reported metrics, the limited knowledge of these caveats, as well as the lack of data and alternatives seem to nudge productivity research into a dependency on them (Bluyssen, 2012; Brutus et al., 2013).

The second part of chapter 3 investigates the value of recollection of productivity in the domain of work from home. I exploit a unique opportunity to look at how (and how accurately) people recall productivity. This situation is especially unique since the experienced productivity has arguably never before changed so rapidly over a short period of time without a change in the core work activities. In the past, productivity changes were often investigated between different groups whose behavior would not be to be independent of their work activities (e.g. the effect of sedentary behavior on productivity; Ishii et al., 2018; Lee et al., 1996) or were simple relatively stable over time (Prasad et al., 2004; Sauermann, 2016). The volatility of the emotional sentiment related and unrelated to productivity during this pandemic period means that accurate recollection is unlikely going to be the “same as before”. But if people are not able to realize that opening a window makes them more satisfied with the office and everything in it, are people able at least to disentangle the past and current sentiment and accurately recollect how productive they were before?

The short answer is no. I ask a sample (N=772) to score their own productivity (and the other four WPSQ factors) during two measurements in June and November 2020. Most notably, during November, I ask these participants to recollect how they had scored themselves in June for all WPSQ factors. Like many memory experts would predict, and similar to all other factors I measure, the recollection of productivity is highly influenced by their current productivity. The recollection of June is generally more strongly, and sometimes even solely, related to the November score instead the actual score for June. For productivity interventions, this implies that self-reported changes do not necessarily reflect an actual change. Asking people how they felt before might provide more (or sometimes only) information about how they feel right now.

1.6 The value of introspection in performance under adverse conditions

The observation that the pandemic has had an effect on recollection does not devaluate self-report completely. Even if the current state has an effect on recollection, and subtle factors such as ventilation have an effect on satisfaction levels, the self-reported effects of big events on productivity as it is happening must at least hold merit. In order to judge the validity of self-reported satisfaction, a controlled environment with manipulation is needed, in which satisfaction, self-reported judgment, and objective performance can be compared.

In Chapter 5, I therefore assess the effect of indoor climate factors on human performance, focusing on the effects of indoor temperature on decision processes. Specifically, I expected heat exposure to influence higher cognitive rational processes negatively, forcing people to rely more on intuitive shortcuts. At the same time, I let people rate their satisfaction with the environment (IEQ specifically) and let them assess the effect it has on their performance. In a laboratory setting, participants (N=257) were exposed to a controlled physical environment with either a hot temperature (28ºC) or a neutral temperature (22ºC), in which a battery of validated tests was conducted. Heat exposure did not lead to a difference in decision quality, but the results did show the limited validity of self-report. A strong gender difference in self-report emerged, such that only men expect that high temperature leads to a significant decline in performance, which does in fact not materialize. For risk, I even show opposites effects: men don't change their actual risk behavior, but state that they are willing to take less risk. Women don't report any difference in willingness to take risks between conditions but actually become significantly more risk-loving. These results confirm our previous doubts about the validity of self-report as a proxy for performance under different indoor climate conditions.

Taken together, chapters 4 and 5 suggest that relocating to work from home could work, as long as we take the physical environment into consideration. Chapters 3 and 5 strongly suggest that the evaluation of this move should not be solely dependent on self-report.

1.7 Social network risks

Whereas working at home could continue for long after the pandemic, our social deprivation was something most of us were eager to get rid of. After a period of lockdown and social isolation, most of us were actively planning to revive our social life as soon as possible. However, even deep into 2022, social interaction did not come without risk. Connecting with a social network could lead to infection, which would inevitably throw you back into physical and social isolation in addition to the obvious risk to your health (Morse et al., 2006; Sun et al., 2020; Ventresca & Aleman, 2013). Nevertheless, for many, total risk avoidance was no longer an option. So how did people decide to meet up with their network?

Unfortunately, not a lot is known about how people estimate risk or spread in networks. And that is surprising since social network assessment is not only important for risk and disease. Innovation, information, and collaboration are key examples of benefits one could 'harvest' from a network (Abraham et al., 2009; Brennecke, 2020; Brennecke & Rank, 2017; Hagen et al., 2018; Pinheiro, 2011; Sun et al., 2020). And in order to 'harvest' (or in the context of the disease: avoid), it is crucial humans understand and estimate the likelihood of how that-which-is-shared-through-through-the-network moves (or spreads) through the network. But grasping the spread through a network is not easy and often requires complex probability computations (Centola & Macy, 2007; Gigerenzer & Gaissmaier, 2015; Pastor-Satorras et al., 2015). Research needed to help us understand is often constrained by the fact that, generally, network assessment deals with multiple layers of uncertainty (Gigerenzer & Gaissmaier, 2015; Hafenbrädl et al., 2016; Simon, 1956). For instance, what qualifies a link for spread? For innovation, should there be a history of collaboration or rivalry, or should all the CEO's children go to the same kindergarten? And if that link is established, what is the likelihood it will materialize? In some conditions, you might be connected to the right people, but still not get invited to parties (e.g. Amabile, 1988; Ventresca & Aleman, 2013).

The risk of COVID-19 infection from a social network during the pandemic does not suffer from many of these classic uncertainties. Everybody is susceptible and for a long time, the likelihood was carefully updated and communicated in the form of the reproduction number, the average amount of people an infected person spreads the virus to (World Health Organization, 2020). Note that the reproduction number does not literally indicate the likelihood of spread, since it is dependent on how many people the infected meets.[22] Nevertheless, never has there been so much information available for a social network problem that applied to so many people.

This cascade of information and the uniquely concrete and identifiable context enables me to look at how people perceive networks in chapter 6. Specifically, I expect that the physical characteristics (such as size and shape) of a network are aiding the complex computations often needed for accurate network spread calculations. For instance, it's possible that individuals perceive some network characteristics as riskier than others by default. The bounded rationality framework predicts that the complexity of probability calculations could be at least partially be substituted by simpler rules of thumb (Sent, 2018). People have the motivation to be accurate, but their constrained mental capacity forces them to (consciously or not) use simpler tools for guestimation (Gigerenzer & Selten, 2002; Simon, 1990). Errors in estimations, in this framework, do not stem from ill intent or lack of motivation, but simply the wrongful application of the right heuristics. More specifically, the majority of the rules-of-thumb perform very well in isolation (e.g., the closer an infection, the bigger the infection risk), but the interaction between multiple conflicting characteristics could lead leading to inaccuracy. So do people use physical network characteristics to estimate spread, and if so, which ones do they rely on the most?

Using a best-worst choice experiment, I let individuals (N= 697) repeatedly rank three varying infected networks, randomly drawn from a set of fourteen, from most risky to least risky. Those rankings enable me to assess the preference between networks, but also the preference of the attributes of those networks. The first result is that the networks are not ranked according to the objective probability of infection risk. That suggests that perceived spread is estimated using other information. I show that humans' processing of risk in networks depends on the

[22] An R of 5 (meaning on average 5 people get infected by an infected person) does not mean that somebody that only sees 2 people will have a higher likelihood to spread than a person that meets 20. They simply influence the R metric differently.

simple characteristics of these networks. Factors such as how close the closest infected contact is to the decider or the ratio of infected to healthy network members have stronger predictive values than the objective probability. Those factors are so valuable for risk perception, that they also predict the network ranking relatively accurately. From this, I conclude that, indeed, the often-complex mental calculation of objective spread in networks seems at least partially substituted by a heuristics-driven approach. These results generalize to the disease context to the extent that policymakers trust blindly the ability of individuals to apply this probability metric objectively. By doing this, this chapter suggests that they are underestimating which other factors individuals actually rely on when connecting to social networks in order to make sense of this abstract probability. But even beyond the pandemic context, I argue that in order to 'harvest' from a network, we must understand how we assess such a network (Burt, 2017; Granovetter, 2018b).

The remainder of this dissertation follows the sequence of the chapters discussed above. After that, chapter 7 summarizes the results into a short discussion, and chapter 8 closes with the societal impact of my findings.

Chapter 2

Avoiding Crowded Places during COVID-19

Common Sense or Complex Strategic Decision?

*[23]

2.1 Introduction

Since the outbreak of COVID-19, countries across the globe have attempted to find ways to contain the rapid spread of the virus. Following a period of strict lockdowns, most countries proceeded towards a policy in which citizens were expected to avoid crowded places, as advised by the WHO (World Health Organization, 2020). Limiting movement to local recreational hotspots as well as (inter)national holiday destinations is considered essential in combating the swift diffusion of COVID-19 infections. Even during the "second wave," avoiding crowded places remains the cornerstone of worldwide policies (National Center for Immunization and Respiratory Diseases (NCIRD), 2020). Following policy advice, however, has proven to be more challenging for the population than initially expected. Popular recreation spots often remain well-visited and shopping centers are almost as crowded as they were a year ago, especially in large cities (BBC News, 2020). Over the Summer of 2020, news and social media showed crowded beaches and partying adolescents almost on a daily basis.

The increase in people visiting crowded places appears irrational from a health perspective, but might be less surprising than expected. Accurately assessing the risk of self-behavior proves to be hard, the urge to recreate seems to grow over time, and the duration of the current situation is testing the limits of human patience and self-control (Huremović, 2019).

---

[23] * This chapter is based on Stroom, Eichholtz, et al. (2021), and co-authored with Piet Eichholtz (Maastricht University) and Nils Kok (Maastricht University).

Moreover, what is considered "crowded?" This uncertainty increases the number of factors and potential outcomes individuals consider (Martínez-Marquina et al., 2019). Whereas recent research discusses theories explaining refusal to comply to COVID-19 restrictions (Demirtaş-Madran, 2021), little to no attention has yet been provided to the thought process that underlies the decision to leave the home, against most policy recommendations, and visit popular recreation areas or crowded shopping streets. Understanding the human thought process from a behavioral perspective, beyond merely labelling behavior to be defiant, will help governments to be more effective in implementing COVID-19-related policies.

This paper investigates the decision of individuals whether or not to avoid crowded places, in a representative sample of the Dutch population, aiming to identify decisive factors underlying this choice. We expect the dependency of the outcome of one's own action on the (unobservable) actions of others to dominate the decision-making process. Therefore, we specifically examine the effect of social context on the decision to visit a crowded place. We hypothesize that providing information on the crowdedness in general will be crucial in the decision of individuals to go out. Specifically, strategic decision making will motivate most people to go out when they expect others to stay home, inevitably leading to escalation and subsequent failure to avoid crowdedness. Similarly, explicit escalation will follow once the general expectations are that most people will go. The aim of this paper is threefold: First, we discuss which decisional processes and conflicts arise due to the ambiguity in the current policy, through the lens of a theoretical framework. Second, using experimental data we demonstrate that (social) context significantly influences the decision-making process of individuals. Finally, we show which personal characteristics have an effect on the decision not to go and how this differs per context. The latter also allows us to draw conclusions on which decisional processes drive the behavior of individuals.

2.2 Theoretical Framework

2.2.1 The Need to Leave

Psychology is unanimous about the inherent human need for social interaction. Baumeister and Leary claim that the need for frequent interactions with others is a necessity for emotional stability (R. F. Baumeister & Leary, 1995). We desire both close individual contact, as well as the ability to function in social groups (Bugental, 2000). Not meeting these requirements leads to invasive negative effects, including, but not limited to physical health and mental well-being (R. F. Baumeister & Leary, 1995). Poor social relationships are estimated to have an effect on mortality similar to smoking 15 cigarettes daily (Holt-Lunstad & Smith, 2012). Recently, a review by Serafini et al. (2020) confirmed the negative impact that frustration, boredom, and disabling loneliness have on (mental) health, specifically following the current COVID-19 pandemic. Social support is one of two main protective factors to avoid mental health issues during this crisis.

The COVID-19 pandemic threatens the ability to meet these basic social needs. This leads to a clear cognitive conflict: people are craving for social contacts, regardless of rapidly rising contaminations with the virus. Using the health belief model framework (HBM; Champion & Skinner, 2008), even without a change in susceptibility or severity of an infection, the downside (i.e. barriers) of staying home slowly start to compete with the benefits. Such a cognitive conflict, better known as cognitive dissonance, can be dealt with in two ways: changing the behavior or changing the reasoning (Festinger, 1957). From a societal perspective, reasoning in favor of keeping distance at all cost, taking no risk, would be preferred. However, the need to socially interact is growing: we observe society-wide violations of the universal policy discouraging social interactions (BBC News, 2020). Going out and being amongst people (albeit within the set regulations) is gaining traction over the safer, more certain option to stay at home to avoid health risks.

2.2.2 Strategic Decision-Making

Acknowledging that the motivation to recreate is strong, the actual decision to "go" or "not to go" to a crowded location depends on the information that is available to the individual at the moment of making the decision. The recommendation to avoid crowded areas is not black or white, and it requires each individual to estimate which spots are considered popular at a given point in time. Although we can assume that every community has a relatively objective view of what is considered a crowded area, the recommendation to avoid these areas implicitly requires an individual to correct for the current situation: how busy will a potentially crowded area be at the moment that I intend to visit it? The degree of crowdedness is determined by the number of people considering to go, their thought processes, and their final decision.

We argue that the seemingly simple choice to visit or avoid a crowded place implicitly involves at least three complex strategic decision-making aspects. First, the choice generally draws parallels with the tragedy of the commons. Hardin describes the tragedy in which a shared yet unregulated common good (in this case, the location for recreation), is spoiled by society because each individual acts according to his or her self-interest, so "depleting" the common good (Hardin, 1968). The similarity lies especially in the fact that collective cooperation would retain the common good, but the individual interest conflicts with the collective maintenance of the good. In this situation, each individual selfishly wants to be in the minority group that visits the recreation area. When too many people act selfishly, the area becomes too crowded and the location no longer meets the "avoid crowded areas" requirements to minimize the spread of COVID-19 infections. In a worst-case scenario, "depletion of the good" could be the closure of the area for recreation, or even reimplementation of a full lockdown.

Second, and more formally, the dependency of each individual's outcome on the choice of the remainder of the population initially resembles the classic game theoretical prisoner's dilemma: going out will lead to a positive outcome if the majority of the population stays away, and only leads to a negative outcome if the majority of the population goes. This dilemma shows us that staying home is not a Nash equilibrium (e.g. an outcome of a decision in which no player has an incentive to deviate from his strategy; Nash, 1950). If everybody stays at home, each individual can improve his or her personal situation by going out. Going out, however, could be considered Nash equilibrium: when everybody is going out, staying at home would not improve somebody's personal situation, when they would be the only person at home. Different from the prisoners dilemma, however, is the Note that we assume that staying at home while everybody else is recreating comes at a (small) disutility, based on the fear of missing out (Przybylski et al., 2013) and not being able to meet the social craving. This makes the decision process oddly circular, and the outcome of the process depends heavily on the moment each individual breaches this circle.

Therefore, third, the decision process to optimize the outcome of the decision concerns *k*-level thinking and cognitive hierarchy theory (Camerer et al., 2004; Stahl, 1993). The core of this theory starts from the premise that hypothetically, all people in strategic games should be able to reason perfectly about their situation, and know that everyone else shares the same capability (e.g. infinitely intelligent). *K*-level theory formalizes this by describing cascades of reasoning levels. The levels refer to the reasoning level someone applies themselves and expects the others to have, or "depth". For instance, the initial level 0 thinkers are considered non-strategic, choosing at will. Level 1 thinkers assume a majority of level 0 thinkers, and will strategically counter the actions of level 0 thinkers' decision. Level 2 thinkers will, in turn, assume a majority of level 1 thinkers, and so forth. At each level, deciders assume that the majority of the group is -*k* level to infinity. In our example, we could hypothetically assume that level 0 thinkers "naively" stay away from a recreation area. As such, level 1 thinkers would come to the conclusion to go as the area will not be crowded. Consequently, level 2 thinkers stay away again, and so forth. Whereas the classical examples of infinite *k*-level thinking would theoretically result in an equilibrium (see for instance the beauty contest example; Nagel, 1995), our context does not. In reality however, the average population's depth of reasoning hardly seems to reach more than two iterations, far from infinite intelligence (Camerer et al., 2004; Ho & Su, 2013). In non-infinite, relatively shallow iterations, *k*-level reasoning could be uses to explain respondent patterns in this context. The implications of the decision to leave home and visit a crowded location during the pandemic are crucial, since citizens likely aim to anticipate the behavior of the majority. When most people are at the same (fairly) low reasoning level, but expect to "outsmart" their fellow citizens (or reason at k+1 level), the chances of an unexpectedly crowded recreation area become very high. Ironically, even when effort is exerted to outsmart the majority and recreate when the majority stays home (thus intending to meet the policy requirements), the implications of cognitive hierarchy theory suggest an "accidental" or implicit escalation of crowdedness.

It is important to note that we approach this policy assessment from a static one-shot perspective. Most game theoretical models also extend to extensive repeated games (often with incomplete information; see for instance Aumann et al., 1995). In that context, belief updating becomes highly relevant. Information can be distilled and adjusted based on the observed outcome of their (and others') previous decisions. Although belief updating is relevant in this context (i.e. witnessing a crowded area whilst expecting it would be empty will likely result in adjustment of a strategy), our aim is to understand if and how the initial expectation of others reasoning and prediction of their behavior influences the first decision to go out or stay. Such insights could prevent early misalignment with personal intentions, minimizing the need for updating and adjustment at a later stage.

2.2.3 Explicit Escalation

In addition to accidental escalation due to the application of wrong strategies by individual citizens, we must also consider explicit escalation, including conscious violation of policy recommendations. In this context, we consider the possibility of the proverbial sheep leaping the ditch: once a large enough group will ignore the policy recommendation, more will automatically follow. These people are, in contrary to the strategic thinkers, no longer intending to *avoid* crowded places. In pandemics this situation is called behavior contamination

(Huremović, 2019). We discuss three types of violations, of which the latter two include cognitive processes that potentially influence the decision to ignore policy recommendations once violations by others are observed.

The first type of explicit violation is based on unrealistic optimism. In contradiction to the latter two types, unrealistic optimism is mostly independent of the behavior of others, as it pertains to the beliefs that the likelihood of something bad happening to you is smaller than it is in reality (Shepperd, Waters, Weinstein, & Klein, 2015). Individuals might violate policy recommendations as a direct consequence of believing that a COVID-19 infection will not happen or harm them. This type of reasoning stems from both the desire to feel good, thus ignoring bad outcomes (Tyler & Rosier, 2009), as well as an overestimation of one's personal characteristics compared to the general population (e.g. being healthier than others; Shepperd, Carroll, Grace, & Terry, 2002)). Although the effect of unrealistic optimism might be smaller for events happening beyond their own control (Klein & Helweg-Larsen, 2002), behavior due to unrealistic optimism is easily distinguishable from other "decision processes" in this situation: individuals will go independent of what other people do or think.

Second, a prevalent view in behavioral science is that these kinds of "deliberate" violations are the result of a loss of self-control or a dominating need to recreate (Huremović, 2019). Boredom and frustration resulting from the ongoing pandemic increases the vulnerability to violate the recommendations (Brooks et al., 2020; Huremović, 2019). Observing others ignoring the recommendations functions as a "broken window": a small violation validates further violations, causing a spread through society (Keizer et al., 2008). This broken window effect, or bad apple effect, is strong even when just a small group of violators is observed (Kerr et al., 2009; Rutte & Wilke, 1992). In this context, seeing others doing something you would also like to do could provide enough of an incentive for citizens to join: why would you stay away if others don't?

Finally, an alternative view explaining why individuals would follow others to crowded places, despite regulations not to do so, involves how people deal with ambiguity. Besides uncertainty about other people's decisions, we also need to consider that people are unsure about the definition of crowded places, or ambiguous regarding the interpretation of the recommendation. Should one take the recommendation as a strict rule, or interpret it more loosely? When ambiguity rises, we tend to use informational social influence to guide our decision (Deutsch & Gerard, 1955). This could lead to contradictions. For example, during the initial loose recommendation to wear face masks in public in the Netherlands, compared to the predominately mandatory use in the rest of Europe, 64% of Dutch citizens were in favor of making face masks mandatory. However, only 17% already wore them at that time (De Hond, 2020). Even when our personal opinion or preference might deviate, in practice we conform to (what we think is) the majority opinion in ambiguous situations (V. L. Allen, 1965). It is crucial to observe from this example that even in a contagious disease pandemic, in which rationally safety is absolutely not in numbers, other people's behavior is still valued in situations of ambiguity. Observing others violating the recommendation to avoid crowded places could therefore be interpreted as the opinion of the majority, and act as information for one's own judgement.

The distinction between the latter two views lies predominately in the underlying intention of the conscious violation. Under the former, the intention can be categorized as ill-intentioned, to the extent that there is no attempt to validate the violation of the recommendation at the start. This does not exclude the possibility that individuals will exhibit post-hoc justification, fabricating reasons why the violation was acceptable or ethical, potentially in response to social disapproval (for instance, after not getting infected with the COVID-19 virus, people could argue that they were correctly assessing the risk ex-ante; (Curley et al., 1986; Haidt, 2001). Under the latter, the intention to deviate from the recommendation originates from confusion. We argue that this behavior reflects the inability to self-assess the ambiguity or uncertainty, leading to herd behavior (Muchnik et al., 2013). Distinguishing between these motivations might be possible by looking at the behavioral response to increasing social violations: for people motivated by ill-intention, going to a crowded place is linearly related to others going; for uncertainty-motivated people, this relationship might only be detrimental when a large enough group signals the "okay" to go. Regardless, however, both motives will inevitably lead to escalation.

2.3 Material and Methods

2.3.1 Participants

We surveyed a panel of 1,048 individuals via Flycatcher, a well-regarded Dutch research organization with access to a high-quality panel used for top research (Bults et al., 2011; Peperkoorn et al., 2020), about their choice whether to go or not to go to a hypothetical recreational hotspot. Our randomly drawn sample from this panel was reimbursed for participation. This sample is heterogeneous in relevant personal characteristics, such as age

(M=43.70, SD=12.52), education, gender (42% male), and occupation.[24] We employ no explicit exclusion criteria, beyond restricting our sample to adults residing in the Netherlands. For an extended overview, see Table 2.1. This research was reviewed and approved by Maastricht University's Ethical Review Committee Inner City Faculties (ERCIC_195_09_06_2020).

Table 2.1 Summary Statistics

		Mean	SD	Minimum	Maximum
Personal Characteristics	Education Category	2.46	.63		3
	Male	42%			
	Age Category	2.18	.74	1	3
	Age	43.78	12.53	18	67
Risk Attitude	General	5.09	1.92	1	10
	Social	5.23	1.95	1	10
	Health	3.87	2.02	1	9
Relatability	Similar	4.06	2.64	1	10
	Imaginable	6.28	2.55	1	10
COVID-19 exposure	Reported Cases	.0004188	.0005787	0	.0041598
	Hospital Admissions	.000108	.0001425	0	.001125
	Deceased	.0000516	.0000736	0	.0005043
	N	927			

Note: Age categories are coded as 1 (18-30), 2 (31-50), and 3 (50+). Risk attitude is measured on a 11-point scale. For Health, a maximum risk score of 10 is never given. Relatability is measured on a 10-point scale. COVID-19 exposure measures are absolute values per 100 inhabitants. In our statistical analysis, these metrics are transformed logarithmically.

2.3.2 Methods

Each respondent is asked to envision the following situation: You live within 20 kilometers of a beach, river, forest, or lake. Under normal circumstances, you (and your household) will seek recreation, cooling and refreshing at this area when temperatures exceed 25 degrees Celsius. You do not have a comparable alternative at home. We ask each participant to decide whether they will visit this area tomorrow, given that it will be 30 degrees Celsius, in five different situations. For the first two situations, the government's recommendation differs: 1) "Stay home", and 2) "Avoid crowded places". For the remaining three conditions, we keep

[24] For an overview of the occupational division in our sample, see Figure 2.2.

the government's recommendation constant ("Avoid crowded places"), but we provide additional information about the situation on the streets: 3) "You see that it is still very quiet on the streets", 4) "You see that the streets are slowly getting busier", and 5) "You do not notice any difference in the degree of crowdedness as compared to last year". We respectively label these levels of context as "Low", "Medium", and "High". All scenarios are presented to the respondents in a randomized order. The high scenario specifically mentions the last year (pre-COVID-19), as we aim to evoke a familiar context which would be similar to all participants. The medium and low context are considered adjustments to that baseline.

We ask each respondent to state whether they will visit the recreation location in each of the scenarios by answering either "yes" or "no". Next, for each randomly presented scenario, we ask participants what percentage of all other respondents they think will answer the previous question with "yes". This percentage provides us with an indication of the expectation that participants have about the behavior of others.

Furthermore, we collect data via the Dutch Bureau of Statistics (CBS) on the local intensity of COVID-19 infections, hospitalizations, and COVID-19 related deaths. COVID-19 exposure is estimated using official government data matched to each individual on geographical location due to the fact that testing was severely limited until six weeks before the experiment. By using public data, we avoid the subjective estimation that people 'might' have had it, influenced by individually factors such as different health beliefs. Using postcode estimation, personal characteristics do not influence the base-rate possibility of exposure to COVID-19 infections[25]. These statistics are matched to each individual in the sample at the four-digit postal code level.

Additionally, we ask the respondents to state their general, social, and health-related risk attitudes on a Likert scale from 0 to 10 (Falk et al., 2016). The risk attitude questionnaire consists of validated questions, one per domain. For example, the general risk attitude question is formulated as follows: "How willing are you generally to take risk". The answer scale for all three questions ranges from "totally not prepared to take any risk" to "very much prepared to take risk". This questionnaire has proven to correlate heaviliy with more extensive and tedious risk attitude measures such as the lottery task. [26][27].

[25] See the limitation section for the discussion of the added value of including health beliefs beyond an indicator for COVID-19 exposure.
[26] For an elaborate overview of the reliability and validation, see Dohmen, Falk, Huffman, Sunde, Schupp, & Wagner (2011).
[27] For an overview of the pairwise correlations, see Appendix Table 2.5.

Although the same recommendation of avoiding crowded places is a COVID-19 policy cornerstone throughout Europe (World Health Organization, 2020), the experienced situational context and timing of our survey is important to ensure external validity. The Dutch government issued an "intelligent lockdown" from March 15th until May 11th 2020. Until June 1st 2020, Dutch citizens were asked to stay home as much as possible. From June onwards, the recommendation to avoid crowded places became the main policy recommendation.[28] Our respondents completed the survey during the first half of July 2020, five to six weeks after the introduction of this recommendation. At this time, the Netherlands had just over 51,000 confirmed cases of COVID-19, almost 12,000 hospitalizations, and just over 6,000 COVID-19 related deaths since the beginning of the outbreak (Statistieken over het Coronavirus en COVID-19, 2021). The timing of our data collection ensures that respondents had ample experience in dealing with the key policy recommendation and that the responses accurately reflected their current behavior. We furthermore consider it important that no new changes in the recommendations were announced at the time, such that the anticipation of new rules, or the signaling of a more liberal approach interfered with the validity of the response.

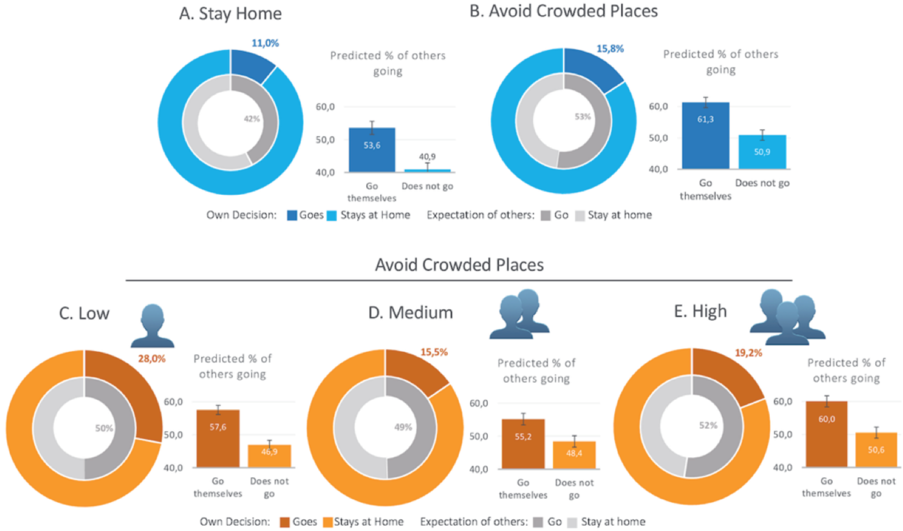
2.4 Results

2.4.1 The effect of context on decision making

Figure 2.1 presents whether or not respondents will visit a crowded area. In all scenarios, the vast majority of the respondents is not planning to go the recreation area. Although this appears encouraging for the policy objective to avoid crowded places, an average of 19% of all respondents across all five scenarios still decide to go.

[28] Note that institution trust dictates the likelihood of adherence of the population to any policy recommendation. Therefore, as background information, the second quarter of 2020 showed the highest institutional trust by the Dutch population is the last 50 quarters (Burgerperspectieven, 2020). For instance, compared to the first quarter of 2020, the trust in the government rose from 51% to 74%.

Figure 2.1 Statistics of intention to visit the crowded place.



Note: The outer ring of the graph shows the percentage of respondents indicating to visit the crowded place, for each context. The inner ring shows the average expected percentage of others to visit the crowded place. The bar graph shows the same expected percentage of others to visit a crowded place, but split by respondents who indicate to go themselves versus respondents who indicate to stay at home. (A) Shows the metrics under the policy "stay home" without any further context. (B) Shows the same metrics but in the condition of "avoid crowded places." (C–E) Show the graphics in this same condition, but each for a different level of crowdedness on the stress (low, medium, and high crowdedness, respectively). For an overview of the difference testing, see Table 2.2 and Appendix Table 2.6.

Panel A shows the percentage of respondents indicating to go to the recreational area when the advice is to "stay home" (10.97%). The inner ring shows the average expected percentage of others to visit the crowded place (42%). Looking at the difference between the recommendation conditions "Stay home" (A) and "Avoid crowded places" (B), we observe a difference of just 5%. Finally, the bar graph shows the same expected percentage of others to visit a crowded place, but split by group of respondents who indicate to go themselves versus people who indicate to stay at home. For instance, for Panel A, people who go themselves predict that on average 53.61% of all other goes (SD = 20.21), whereas the people who stay at home predict only 40.92% to go (SD = 20.51). The difference between these two groups is statistically significant (see AppendixTable 2.6; z = -6.07, $p < .001$).

When we add context about the level of crowdedness on the streets, we observe an additional increase in the number of respondents intending to leave the home. It is noticeable that providing a *clear* context about the crowdedness on the streets, regardless whether this is low (C) or high (E), causes a steep increase compared to the middle condition (D) and even no context (B). Panel A of Table 2.2 shows the results of a series of proportion test comparing the proportions per condition. It shows that likelihood of going out does not differ significantly between no context (B) and the middle condition (D) (diff = .044, p=.81). Both the low (C) and the high (E) condition differ significantly from both no context (B; diff = .121, p<.000, and diff = .034, p<.05, respectively) and the middle condition (D; diff = -.125, p<.000, and diff = .037, p<.05, respectively). Respondents are more likely to go to the area of recreation when they expect it to be quiet (overall most likely, even compared to the second most likely condition: high (E); diff = .088, p<.000). This is in line with both the official policy recommendation as well as strategic thinking. Respondents are also more inclined to visit a popular area when they have reason to believe that it will be crowded at this location. This is directly opposite to the official policy recommendation, and not in line with game-theoretical predictions. This preliminary result suggests that respondents' strategic thinking (in the low context) as well as social norms (in the high context) play a role in their decision whether to go, or not.

Table 2.2 Statistical testing of the difference between conditions: going versus not going

	No context	Low	Medium
Panel A – Proportion of going			
Low	.121***	0	
	(6.71)	(-)	
Medium	-.004	-.125***	0
	(-0.24)	(-6.94)	(-)
High	.0338*	-.088***	.037*
	(2.01)	(-4.73)	(2.25)
N	1,048	1,048	1,048
Panel B – Predicated percentage others' going			
Low	-2.63***	0	
	(-5.52)	(-)	
Medium	-3.10***	-0.45	0
	(-6.73)	(-0.91)	(-)
High	-0.16	2.48***	2.94***
	(-0.34)	(4.68)	(7.34)
N	1,048	1,048	1,048

Note: For both panels, the score is constructed such that the mean value of the row conditions is subtracted from the column conditions. Panel A shows the difference in proportions (proportion test) of people going out under different conditions. The outcome variables are binary such that 0 = not going and 1 = going. For example: people in the low context go on average 12.1% (.280 minus .158, respectively) more often as compared to the no-context condition. Z-statistics in parentheses. Panel B shows the difference in predicted percentage (scored between 0 and 100) that people expect others to go. For example: people in the high condition expect that others

We then investigate the estimation that respondents make about other's behavior (Panel B, Table 2.2) using Wilcoxon rank-sum tests. We observe that respondents substantially and consistently overestimate the number of other people intending to go. Respondents expect, on average across all scenarios, that roughly 50% will decide to leave the home and recreate.

Furthermore, the predicted percentages do significantly change between scenarios. We see significant changes in the prediction of other people's behavior, indicative of the motivation of individuals to go themselves. For instance, introducing the "low crowdedness" context (C) compared to no context (B) almost doubles the number of respondents planning to go to the area of recreation (proportional increase of 12.1 percentage points, z=6.71, p<.000), when the expected percentages of others going drops with 2.63% (z = 4.68, p<.000). Interestingly, moving from no context (B) or medium context (D) to high context (E), increases the proportion of people going with roughly 3.5 percentage points (3.4%, z=2.01, p=.04; 3.7%, z=2.25, p=.02, respectively), when also the prediction of others going increases (significant only for medium to high context; 2.94%, z=7.34, p<.001). In general, introducing low context information increases going out whilst the expected percentage of others going out drops. Introducing high context information increases the likelihood of going out in conjunction with an increase in the expected percentage of others going out. However, the limited absolute value changes in the expectations about others indicate that the changes in one's individual decision to go are not fully reflected in the prediction of other citizens' behavior. In general, the expectations about others' behavior are a lot more negative than one's own behavior, and more negative than the behavior of the collective.

There is also a relationship between going out yourself and the expectations about others. Non-parametric Wilcoxon rank sum tests show consistently that, regardless of the scenario, respondents indicating a willingness to recreate themselves also predict a significantly higher number of people to make the same decision, compared to respondents indicating to stay home (all are significant for p <.001; for an overview of these statistics, see Appendix Table 2.6). The prediction is significantly correlated with citizens' own decision to go: for each percentage point increase in the prediction that others will go, the marginal effect of going themselves increases with an average of 0.3% (results are not presented in the table: ranging from 0.2% to 0.4%, p <.001 throughout all contexts).

2.4.2 Predictors

A key question is which factors are decisive for choosing to leave the home for recreation in each of these conditions. Table 2.3 investigates the role of personal characteristics in the choice for recreation per condition using a logit regression. The results show that education plays a key role in the decision to go, despite the regulation. The low education group turns out to be most likely to abide by the rules. The middle-educated category (post-secondary vocational degree, undergraduate education, or higher level of high school) are generally more inclined to go, compared to the low education group (post-secondary vocational education or lower-level high school). The most highly educated respondents (undergraduate degree or higher) indicate an even higher willingness to go. The effect of education is most profound in the low (for middle education the marginal effect is 12.9%, z = 2.43, p = .015; for high education the marginal effect is 17.9%, z = 3.29, p =.001) and high context conditions (middle: 10.6%, z = 2.41, p = .016 and high: 11.0%, z = 2.51, p =.012, respectively). In the "medium" condition, we find no effect of education.

Table 2.3 Logit regressions: respondent characteristics and decision to go.

		No context	Low	Medium	High
	Female	0.984	1.006	0.963	0.995
		(-0.65)	(0.20)	(-1.57)	(-0.20)
Education	Middle	1.073	1.129*	1.060	1.106*
		(1.85)	(2.43)	(1.43)	(2.41)
	High	1.095*	1.179**	1.069	1.110*
		(2.27)	(3.29)	(1.61)	(2.51)
Age	31 t/m 50	0.967	0.903*	0.990	0.991
		(-0.92)	(-2.27)	(-0.29)	(-0.25)
	Above 50	0.907*	0.791***	0.926*	0.919*
		(-2.57)	(-5.12)	(-2.12)	(-2.16)
Risico attitude	General	1.008	0.983	1.004	1.002
		(1.01)	(-1.76)	(0.55)	(0.21)
	Social	0.995	0.997	0.989	0.991
		(-0.70)	(-0.29)	(-1.61)	(-1.22)
	Health	1.016*	1.048***	1.020**	1.023**
		(2.57)	(5.54)	(2.96)	(3.16)
	Chi^2	33.63	70.12	29.26	26.36
	N	964	964	964	964

Note: Education is relative to the baseline category "Lower education" and age is relative to the baseline category "30 years or younger." z-statistics in parentheses. Standard errors are clustered at the individual respondent level. * p < 0.05, ** p < 0.01, *** p < 0.001

We also observe an effect for age, but not for gender. The effect for age is negative across all contexts. In the low crowdedness context, both age brackets have a significantly negative marginal effect (-9.7%, z = -2.27, p = .023, and -20.9%, z = -5.12, p < .001, respectively), whereas for all other contexts we observe older respondents (50+) to be less likely to visit the recreation location, compared to the 30 year and younger category. Interestingly, the impact of personal characteristics seems to diminish when the streets are getting busier: in the highly crowded context, both the significance as well as the strength of the effects of education and age decrease as compared to the "low" context.

The general and social risk attitudes do not have a significant influence on the decision of respondents. The degree to which respondents are willing to take risk with their own health, however, is important throughout all contexts. For each incremental increase of willingness to take risk on this domain, the probability that a respondent will go increases with 1.6% ($z = 2.57$, $p = .01$) to 4.8% ($z = 5.54$, $p < .001$) per context. This result implies that the decision to go depends more on respondent's own health considerations than on the fear to contaminate others.

2.4.3 Additional Explanatory Variables

2.4.3.1 Similarity and Imaginability

The hypothetical nature of self-reported vignette studies negatively affects their validity compared to actual behavioral measures (this is also referred to as the intention-behavior gap; Sheeran & Webb, 2016). The decision to go and visit a crowded place on a hot summer day will be influenced by the degree to which each respondent in our sample can relate to this specific scenario. For instance, a person living in a city center without a garden will likely better understand the motivation to go out of the house as compared to a person living in a rural area with big garden. To test whether these location-dependent characteristics influence the decision to go, we measure two additional indicators: level of similarity (e.g., to what extent the situation mimics their own situation) and the level of imaginability (e.g., to what extent are respondents able to imagine being in such a situation). For a summary of these metrics in our sample, see Appendix Table 2.5.

We find that similarity increases the likelihood of visiting a crowded place. Panel A of Table 2.4 shows that for each increase on a similarity scale from 1 to 10, the marginal increase of going out ranges between 2.3% and 1.5% depending on the context (no context: $z = 5.25$, $p < .001$ and high context: $z = -3.07$, $p < .01$, respectively). Beyond similarity, imaginability increases the probability of going out in the low context (1.8%, $z = 2.49$, $p < .05$) and high context (1.2%, $z = 2.03$, $p < .05$). In sum, both the similarity and imaginability of the situation increases the probability of visiting the recreation area, in most contexts.

Table 2.4 Logit Regressions: Location Dependent Characteristics and Decision to Go

	No context	Low	Medium	High
Panel A. Relatability				
Similar	1.024***	1.019**	1.023***	1.015**
	(5.25)	(3.22)	(4.90)	(3.07)
Imaginable	1.004	1.018*	0.999	1.012*
	(0.80)	(2.49)	(-0.19)	(2.03)
Chi²	33.63	70.12	29.26	26.36
Controls	Yes	Yes	Yes	Yes
N	964	964	964	964
Panel B. COVID-19 Exposure				
Reported Cases	1.023	1.076	1.072	1.035
	(0.61)	(1.51)	(1.80)	(0.80)
Hospital Admissions	0.973	0.972	0.948	0.963
	(-0.79)	(-0.63)	(-1.51)	(-0.96)
Deceased	1.010	0.968	0.997	1.012
	(0.43)	(-1.10)	(-0.13)	(0.46)
Chi²	32.26	62.90	29.09	28.19
Controls	Yes	Yes	Yes	Yes
N	840	840	840	840

Note: Panel A shows the marginal effect of the reliability measures on the decision to go. Panel B shows the marginal effect of COVID-19 different exposure indicators, using postal codes, on the decision to go. The measures are per 100 inhabitants, and transformed to natural logarithm due to a highly skewed distribution. Note that sample B consists of a smaller sample due to missing values in the COVID-19 database. Both panels are controlling for all personal characteristics presented in Appendix Table 2.7: education, gender, age, and risk attitude. z-statistics in parentheses. Standard errors are clustered at the individual respondent level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

2.4.3.2 COVID-19 exposure

In order to generalize our results to other situations, and to show that policy and context drive the behavioral intensions that we observe, we assess the impact of COVID-19 exposure on the decision of our respondents to go. It is plausible that the survey participants experience the context we present to them in the light of their own experience of the COVID-19 threat. In order to investigate the robustness of our findings, we match all individual respondents to COVID-19 metrics that are publicly available through the Dutch Ministry of Public Health, using respondent postal codes (RIVM, 2020). Specifically, we standardize reported COVID-19 cases, hospital admissions, and COVID-19-related deaths such that for each postal code the value shows the ratio per 100 inhabitants.

Panel B of Table 2.4 shows the effects of local COVID-19 metrics on the decision to go, for each context. Due to the skewedness of all the metrics, we transformed the metrics using a natural logarithm. First, we find a marginally significant impact of the number of hospital admissions on the likelihood of going out at the medium level of context (-5,2%, $z =-1.51$, $p<.1$). For all other levels, the number of hospital admissions and COVID-19-related deaths do not have an effect on the likelihood of going out. For the number of reported cases we find a marginally significant trend at the 10% significance level, having the opposite effect. Specifically, a larger number of reported cases suggests a higher likelihood of going out, only for the low and medium context, ranging from 7.6 to 7.2% increased likelihood (medium context: $z = 1.51$, $p < .1$ and low context: $z = 1.80$, $p < .1$, respectively). [29]

In summary, the influence of local COVID-19 exposure on our results, based on publicly available COVID-19 data, is weak and inconclusive. We observe an increased trend to recreate when there are more reported cases in the respondent's postal code. However, this correlation could also be reversed in causality: more cases are reported because people tend to go and recreate. On the other hand, we find a comparable yet opposite likelihood of going for the local COVID-19 exposure of hospital admissions. The effects are concentrated exclusively in the "low" and in the "medium" condition, and they are only marginally significant.

Overall, given that both robustness analyses have an effect on the decision to go, we also added COVID-19 exposure measures, as well as the similarity and imaginability measures, as controls in the main regression of Table 2.3 (see Appendix Table 2.7). We observe only minor significance changes and no noteworthy changes in interpretation or direction of our previously discussed main results.

2.5 Discussion

2.5.1 Conclusion

Public health policies to contain COVID-19 infections are under heavy scrutiny. An important pillar of public policies in almost any country is the recommendation to "avoid crowded places." This appears to be a straightforward message, but in reality, it is not, since it inevitably introduces considerations of other people's expected actions in citizens' own decision-making process. Although the results in this paper suggest that the majority of citizens adhere to the policy recommendation[30], the results also suggest that people are implicitly forced to make a correct estimation of the situation outside. This is not trivial to each individual. The results not

[29] The lack of significant effect of the exposure to COVID-19 could be due to a discrepancy between the official numbers and the perceived exposure by each individual. We do not suggest that the perceived exposure would be a more accurate COVID-19 exposure metric, but do acknowledge that this subject perspective (related to health beliefs) could be a relevant explanatory factor in our results on its own. See the limitation section for a further discussion on this topic.

[30] In the limitation section, we discuss some important psychological factors that could potentially explain the proportion of individuals who are not affected by the social context, and will always stay home.

only show that a vast majority of respondents is unable to make an accurate estimation about others' behavior, but also that a wrong estimation could lead to a worsened outcome.

In line with the theoretical framework, providing information regarding the situation outside initially leads to a rational choice (e.g., when it is calm, the majority intends to go, and when it is reportedly getting crowded, more respondents intend to stay home). The strategic decision underpinning is most clearly illustrated when moving from "no" context to "low" context: a steep increase of people who go themselves, combined with a significant decrease in the expectation of others to go. However, once people know that it gets even more crowded outside ("medium" to "high"), respondents indicate a greater willingness to go out, combined with an increase in expectations about others going, possibly leading to an escalation in crowdedness. These observations seem to indicate behavior contamination (Huremović, 2019): the stronger the expectation that others will go, the more likely it is that people will go themselves (Keizer et al., 2008). Our results suggest this latter "explicit" violation of the public health regulation is more likely a result of using social cues for ambiguity management than a bad-apple effect. Comparing the behavioral trend from the "low" to "medium" and finally "high" context, we see that moving to more ambiguity (medium crowdedness context) leads to fewer people going (e.g., providing no context is almost identical to the medium context, strengthening the ambiguous interpretation of the medium context). Since we do not observe a linear increase in violation over intensifying crowdedness contexts, but a parabolic relation, we believe it is likely that we witness the social context as informative to behavior, instead of provoking "violating" behavior. Overall, both theoretical predications are supported: strategic decision making seems to motivate people to go out when they expect others to stay home, whereas explicit escalation follows once the general expectations are that more people will go out.

The heterogeneous effects of multiple predictors on the decision to go gives crucial hints on the motivation and underlying thought process per context. A key indicator is the effect of education on the low and high context suggests that educational background is more important in the rational or strategic (low context) decision, than in the escalation (high context). Thus, we conclude that in the low context situation, highly educated people act strategically and in the medium context the social norm is leading in coping with the ambiguity. In the high context, social norms lead to escalation. Second, overall, the willingness to take risk in the health domain is an important predictor to go out: the higher the willingness to take health risks, the higher the likelihood of going out. Interestingly, this effect is strongest in the low context condition. The marginal effect of the willingness to take risk in the health domain is almost double compared to the other conditions. We observe the same for age: older individuals are less likely to go out in general, but the effect is almost twice as big in the low context condition compared to all other conditions.[31] Although our results do not imply causality, and must therefore be interpreted with caution, they are not contradicting our previous conclusion: in the low context, a strategic decision process underlines the decision to go. Education, health, and age weigh heavily in the ultimate decision. These factors weigh less strongly in the "high context" condition, where the decision to go is rather motivated by behavior contagion instead of individual considerations. In other words, in a "low context" situation, people decide themselves, in "high context", others (at least partially) decide. Specifically, in line with the Health Belief Model (HBM; Champion & Skinner, 2008) it is likely that perceived susceptibility and severity of the infection are influenced by the social context. Seeing others go out, might signal that others estimate the severity lower than they themselves do, lowering the motivation to stay inside (leading to behavior contagion).

The context that is given to people in their decision-making process is thus detrimental, but does not have a uniformly positive effect. Additional relevant factors such as willingness to take risk with one's own health and the similarity to one's own situation all increase the likelihood to visit crowded places.

It is also evident that people underreact to the behavior of others. In general, we observe incorrect pessimism about other people's behavior: across all conditions, people expect far more people to go than the collective intention to do so. However, individuals also underestimate the effect context has on others, even when it has a profound effect on our own behavior. In other words: when the context influences people to go, people underestimate the increase in crowdedness as a result of other people making the same judgement due to the same change in context. This causes an escalation in the "low" context: Although the crowdedness context signals a quiet situation at the location of recreation, people do not take into consideration that the majority will come to the same conclusion. As such, our findings result in the somewhat paradoxical prediction that it will be busiest in the low crowdedness context.

In conclusion, the main aim of this paper pertains to assessing the impact of an ambiguous policy to "avoid crowded areas", leaving individuals to form expectations about the

[31] It is important to note that the "low context" condition has 50 to 100% more people going out compared to the other context conditions. The strength of the significance and coefficient in a logit regression is influenced by the total amount that go out in that context compared to other coefficients. However, although the significance of the effect could be more easily detected, the magnitude of the coefficient should be less affected.

level of crowdedness themselves. without guidance on which information they can use to come to this assessment.[32] We show that, providing individuals with an ad-hoc proxy for crowdedness of which the informational value is unclear, leads to suboptimal yet predictable thought processes and decisions. Specifically, we show that a considerable number of people think they are strategically avoiding crowded places when it is quiet outside, and follow the herd when it is busy.

2.5.2 Limitations

We strive to identify how the current (Dutch) COVID-19 policy recommendations, combined with limited information availability, influences behavior of individuals. In doing so, we intentionally strike a balance between a rigid experimentally controlled design, and elicitation of real-life ambiguity that closely reflects the current situation that individuals find themselves in. Loosening the experimental controls often comes at the cost of increasing the likelihood of omitted variables. Below, we discuss three main limitations of this study.

First, to achieve real-life ambiguity, our experiment is intentionally ambiguous in two dimensions: the location of recreation and the level of crowdedness. The first ambiguity increases the probability that the participant empathizes with the hypothetical situation. Specifying the location would surely have increased uniformity in beliefs about the expectations of crowdedness, travelling factors, or density of the location (e.g., how crowed is a beach compared to a forest or city center?). We acknowledge that omitted variables directly related to the preferred location might influencing the decision. However, keeping the location as a general category increases the likelihood that participants are able to envision themselves in this hypothetical location, regardless of their personal preference. This means our results can be generalized. Indeed, respondents in the sample their ability to envision themselves in this situation (average imaginability score of more than 6 out of 10), even though respondents might not necessarily be in this situation (average similarity score is only 4 out of 10). The second ambiguity is on the degree of "crowdedness." This is not stated as an objective measure, but as a subjective experience that depends on the interpretation of the participant. For example, "the streets are slowly getting busier" aims to elicit a general tendency of increasing crowdedness in the community, but could be influenced by the literal interpretation of what the individual considers "the streets" as well as "getting busier." Moreover, we consider these conditions to be

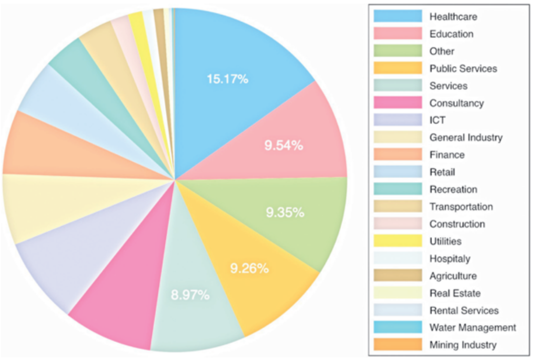
[32] Note that we are explicitly not manipulating communication or policy recommendations.

at least ordinal in our interpretation, but the proportional distance between these levels can only be assumed. We are therefore unable to exclude that, in both dimensions, the interpretation of the ambiguity may lead to other reasoning and thus other behavior than we anticipate. However, note that these ambiguities are present in real-life decisions as well. We argue that the value of generalizability (at least partially) compensates for these potential omitted influences.

Second, we take a wide variety of individual factors and traits into consideration, but must acknowledge that additional personal beliefs and traits might matter as well. Most profoundly, we explicitly do not discuss motives for individuals to stay at home throughout all conditions and contexts. For instance, Jeong et al. (2016) mentions the most frequent reaction to a pandemic to be the uncontrollable fear for infection. Individuals who were exposed to infection are more likely to develop worries about their own health and infecting others. Especially pregnant women and parents with children are likely to develop such fears (Braunack-Mayer et al., 2013). By focusing on individuals who consider to go, we neglect motives to not go at all. Seeing that the majority in our sample chooses not to go out at all, we feel strongly that psychological factors such as pervasive anxiety and uncontrolled fear (Serafini et al., 2020), as well as individual self-efficacy and perceived benefits of staying at home (Health Belief Model; Champion & Skinner, 2008), are key drivers for this behavior. However, this paper does not focus on the decision to go at the extensive margin, and is therefore unable to explain key drivers not to go at all, regardless of social context. More extensive research should focus on including and identifying the crucial factors determining the absolute choice to stay home.

Moreover, although we include risk aversion (in multiple relevant domains), demographic differences, and personal exposure to COVID-19 in our analysis, we do not include personality traits. We also need to acknowledge that, although we strived to approximate personal exposure to COVID-19, we are unable to identify frontline healthcare workers who are exposed to a uniquely intense level of exposure incomparable to private life exposure. Note that Figure 2.2 shows that our sample holds over 15% healthcare workers, but we are unable to distinguish between frontline COVID-19 workers and healthcare workers for which exposure is comparable other occupations (e.g., massage therapist, dentists, or physical therapists). Finally, we expect that people with a garden (or perhaps even a balcony) might find the need to recreate outdoors significantly less acute as compared to (large) families in apartments without such amenities. We specifically ask the respondents to consider a situation in which the area of recreation is the only available means of recreation, but we cannot exclude the possibility that other individual differences influence our results.

Figure 2.2 Distribution of occupations.



Note: The graph includes the percentage of respondents for the five largest groups of professions in our sample, making up for more than half our sample. Note that healthcare professions in our Dutch sample include "well-being" ("zorg en welzijn"), which is a broader category than purely healthcare professionals. This also includes massage therapists and physiotherapist, for instance.

Finally, we frame our experiment as a one-shot game even though in real-life, people are able to update their information. Information about traffic jams, live news coverage of popular spots, and even witnessing crowdedness themselves once they are on the road will potentially change behavior. This includes information from past days (e.g. media coverage of previous hot days), current events (e.g. social media coverage of friends and family), and future updating (e.g. once traveling, seeing others on the streets). For some people, this information will influence their decision on the day itself, for others their commitment to their initial decision will be less easily swayed. However, we note that we do not argue that our key take-away is that all popular locations will inevitably end up crowded due to the ambiguous policy. The main result of our paper is that this policy combined with no clear and updated information of the behavior of other participants (e.g., state of the recreation spot) leads to an unintended suboptimal group decision following an (seemingly) optimal individual decision. Without correct information or information updating, this could lead to an escalation of crowdedness.

2.5.3 Implications

The COVID-19 pandemic demands significant self-control from society to stay home. The recommendation to avoid crowded places creates a sense of freedom and offers the possibility to act dynamically given the circumstances. The definition of this policy advice, however, also offers freedom in interpretation. Consequently, the freedom is implicitly asking more from the population than it initially seems. “Use your common sense” is often the accompanied advice, but our results show that more and better information concerning the context is essential to make an optimal decision.

The results of this research are not predominately pessimistic. Besides the fact that the majority of respondents indicates to stay home, we also identify a strong inclination to avoid crowded places. Only after feeling that nobody stays home any longer are people legitimizing their own violation of the recommendation. Furthermore, the existing pessimism that society has regarding the behavior of others could lead to an escalation of the situation. Providing up-to-date information could be detrimental for an accurate estimation of the situation. This information could reinforce and stimulate positive behavior. Both going out as well as staying at home are rational and ethical choices. It is, however, the relevant context that determines whether going or staying leads to a rational decision, or escalation. Without this information, the outcome of a decision will remain uncertain.

Additionally, discouraging unwanted behavior should be tailored to the individuals who are more inclined to ignore the policy recommendation. Young as well as highly educated people are less sensitive to policy recommendations in the calmer contexts, and should thus be discouraged accordingly. They draw valid conclusions, but do not seem to be aware of the potential harmful consequence when a large part of society independently reasons in the same way. Here too, facilitating relevant information could offer a solution, and avoid escalation. Moreover, seeing the violations of this policy in age brackets could spark the discussion of monitoring ‘youth hotspots’ more than other hotspots. If it would turn out that the young remain insensitive to this recommendation even after our suggested enhancement of information, differentiation in monitoring locations could be an effective detergent method policymakers should consider before relapsing to the most restricting policy to ‘stay at home’ for all. However, at this point the interaction between punishment, monitoring, and information provision remains speculative without further examination.

Finally, the risk profile of each individual could offer a potential policy approach. Finding that the risk attitude regarding citizens’ own health plays a key role in their decision to go or stay home, suggests that campaigns emphasizing and educating people about their own health risk could improve the collective behavior of society.

2.6 Appendix

Table 2.5 Pairwise Correlations of Independent Variables

Variables	(1) Education	(2) Age	(3) Similar Relatability	(4) Imaginable Relatability	(5) Reported Cases	(6) Hospital Admissions	(7) Deceased	(8) General Risk Attitude	(9) Social Risk Attitude
(1) Education	1.000								
(2) Age	-0.343***	1.000							
(3) Similar Relatability	0.079	-0.162***	1.000						
(4)Imaginable Relatability	0.165***	-0.173***	0.506***	1.000					
(5) Reported Cases	0.068	-0.046	-0.052	0.017	1.000				
(6) Hospital Admissions	0.083	-0.078	-0.050	0.006	0.958***	1.000			
(7) Deceased	0.089	-0.056	-0.034	0.013	0.930***	0.921*	1.000		
(8) General Risk Attitude	0.068	-0.013	0.124**	0.075	0.004	-0.023	0.017	1.000	
(9) Social Risk Attitude	0.071	0.052	0.049	0.076	0.022	-0.030	0.022	0.478***	1.000
(10) Health Risk Attitude	0.020	0.013	0.103	0.032	0.030	-0.008	0.033	0.508***	0.372***

Note: for this table, education and age are not transformed to categories. Education is on a 0 to 11 scale. All COVID-19 exposure (5-7) measures are stated per 100 inhabitants, and transformed to natural logarithm due to the skewed nature of the distributions. Note that the highest correlated factors were also included stepwise into the main model, to check for collinearity issues. **** $p<0.001$, ** $p<0.01$, * $p<0.05$*

Table 2.6 Statistics on the Expectations of Others Going Within Each Condition

Predicted ratios of others' going		Total	I will go	I will not go	*p*-value
Stay at Home	No context	42.31 (20.85)	40.92 (20.51)	53.61 (20.21)	.00***
Avoid Crowded Places	No context	52.54 (19.93)	50.89 (19.78)	61.32 (18.41)	.00***
	Low	49.90 (22.25)	46.93 (21.08)	57.58 (23.34)	.00***
	Medium	49.44 (20.55)	48.40 (20.05)	55.19 (22.29)	.00***
	High	52.38 (21.79)	50.57 (21.26)	60.03 (22.38)	.00***
	N	1048			

Note: First column shows the overall average predicted percentage of other's going. The latter columns show the same statistic, split depending on the participants going themselves ("I will go") versus staying home ("I will not go"). The size of these subgroups fluctuates per condition and context. Standard deviation in brackets. P-value based on non-parametric ranksum test. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table 2.7 Integrated Logit Regression: Personal and Location Dependent Characteristics and Decision to Go

		No context	Low	Medium	High
Education	Middle	1.072	1.126*	1.063	1.105*
		(1.88)	(2.23)	(1.47)	(2.30)
	High	1.084*	1.160**	1.056	1.096*
		(2.13)	(2.85)	(1.32)	(2.19)
	Female	1.002	1.024	0.975	1.013
		(0.09)	(0.70)	(-1.03)	(0.46)
Age	31 - 50	0.983	0.926	1.006	1.020
		(-0.44)	(-1.58)	(0.17)	(0.53)
	Above 50	0.922*	0.812***	0.938	0.945
		(-2.06)	(-4.23)	(-1.73)	(-1.47)
Risk Attitude	General	1.004	0.981	1.003	0.999
		(0.45)	(-1.81)	(0.35)	(-0.10)
	Social	0.996	0.996	0.989	0.988
		(-0.62)	(-0.49)	(-1.51)	(-1.58)
	Health	1.015*	1.044***	1.015*	1.023**
		(2.33)	(4.85)	(2.37)	(3.14)
Relatability	Similar	1.024***	1.017**	1.023***	1.016**
		(4.90)	(2.63)	(4.70)	(3.00)
	Imaginable	1.002	1.014	0.998	1.010
		(0.39)	(1.84)	(-0.35)	(1.56)
COVID-19 exposure	Reported Cases	1.023	1.074	1.072	1.032
		(0.64)	(1.47)	(1.85)	(0.75)
	Hospital admission	0.984	0.977	0.957	0.970
		(-0.51)	(-0.52)	(-1.28)	(-0.79)
	Deceased	1.005	0.968	0.992	1.010
		(0.24)	(-1.10)	(-0.33)	(0.40)
	Chi^2	62.22	69.83	50.10	41.58
	N	840	840	840	840

Note: Education is relative to the baseline category 'Lower education' and age is relative to the baseline category '30 years or younger'. All COVID-19 exposure measures are stated per 100 inhabitants, and transformed to natural logarithm due to the skewed nature of the distributions. The sample is smaller compared to table 3 due to missing values in the COVID-19 exposure data. z-statistics in parentheses. Standard errors are clustered at the individual respondent level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$..

Chapter 3

The Work Productivity and Stress Questionnaire (WPSQ) and Recall Bias[33*]

3.1 Introduction

Productivity is the pillar of work success and satisfaction. Being successful at work is synonymous with being productive, and especially for employers, high productivity in turn translates to better firm performance (e.g. Krekel et al., 2019; Palia & Lichtenberg, 1999). Not only is a happy employee a productive employee, but the reverse is found to be true as well (e.g. Mamiseishvili & Rosser, 2011; McNeese-Smith, 1996; Miller & Monge, 1986). Productivity (and productivity change) is therefore a potent indicator: high productivity signals good performance or a successful company policy change, low(er) productivity might indicate that action needs to be taken, or policy changes should be reconsidered.

Inferring the stability of, or change in, productivity requires accurate measurement, yet measuring productivity is elusively challenging. When the outcome is either abstract, not standardized, or not easily observed labor, directly measuring the output is not representative (in comparison to manual labour, for instance; Bartelsman & Doms, 2000; Singh et al., 2000; Syverson, 2011). For example, high-skilled workers do not produce the same outcome on a fixed basis, and changing projects and deadlines, as well as evaluation of the final result might differ within and between workers. As a result, most research focusing on high-skilled desk workers relies heavily on self-reported measures of productivity (Bloom & Van Reenen, 2007; Del Gatto et al., 2011; Färe et al., 1998; Gidwani & Dangayach, 2017; Singh et al., 2000; Skirbekk, 2004; Syverson, 2011). This paper sets out to provide versatile, self-report-based, multidimensional productivity and stress questionnaire to approximate high-skilled labor productivity. I reconstruct and validate a Work Productivity and Stress Questionnaire that fits items to a large

[33] *This is a single author chapter

sample of homeworkers. For a metric so pivotal in performance research, this productivity questionnaire aims to be better-identified than its current alternatives, yet still easy-to-apply. The resulting tool can be used as an alternative when objective measures of productivity are obscured or unavailable.

In addition, I explore whether retrospective reports of a self-reported questionnaire remain valuable in the productivity domain. Retrospective memory limitations have repeatedly been shown to decrease the inaccuracy of recollection of past autobiographical sentiments. For instance, people unintentionally let their current feelings influence their retrospection (recall bias; Coughlin, 1990; Raphael, 1987). I utilize this questionnaire to explore the topical pandemic-driven working-from-home shift. The volatility in productivity during that pandemic additionally enables the identification of recall bias in the work and productivity context, previously unexplored.

In this paper, I first redevelop the Health and Work Questionnaire (HWQ) questionnaire originally developed by Shikiar et al. (2001) and assess its performance in a work-from-home context. I apply a principal component analysis on the items and find different factors that fit the data better. I identify, label, and discuss the internal consistency, reliability, and validity of these factors and explore the performance compared to the original questionnaire's factors. The results of the questionnaire validation show that the new principal component analysis outperforms the original questionnaire factors. The restructured questionnaire, which I name the Work Productivity and Stress Questionnaire (WPSQ), identifies five factors: productivity, productivity by others, peer relations, nonwork satisfaction, and stress and irritability. These factors show a clear improvement over the original factors. For instance, the internal consistency of the factors in the WPSQ does not fall below 0.73 (Cronbach's alpha), improving from alpha's often dropping below 0.65 in the original scale. Moreover, the productivity factor more often correlates to relevant single-scale items (both divergent as well as convergent), and the factor of stress and irritability is better correlated with items approximating absenteeism. The latter was previously mentioned as a caveat of the HWQ. Using the repeated-measure, I show that the factors are also consistent for different measurements and reliable over time.

Second, I explore the recollection accuracy over time using within-subject repeated measures. Following a short literature introduction on recollection accuracy and memory pitfalls, I identify a consistent recall bias and explore possible alternative explanations for this phenomenon. The results of the within-subject repeated measure uncover a consistent recall bias. This bias is often overlooked in the productivity domain, yet inherent to self-reported (autobiographical) recollection of questionnaire scores over time. I show that, although the questionnaire WPSQ is internally consistent, the retrospective scores themselves are heavily influenced by a recall bias: the current state colors the recollection of the past. With a regression approach, I debunk the initial observation that suggested that increased current scores increase accuracy.

The contribution of this paper is twofold. First, I develop an improved tool to estimate productivity through self-report. The growing popularity of working away from the office underlines the importance of monitoring productivity and work satisfaction in alternative ways. The multidimensional factors of this easy-to-administer tool extend the assessment beyond productivity, where multiple relevant domains provide more insights into the construction of underlying productivity and work satisfaction. Second, identifying a recall bias in all factors during a highly volatile period warns us to interpret self-reported past productivity with great caution. Without acknowledging this bias, an evaluation of the impact of the first months of the pandemic on productivity could be severely misjudged.

The remainder of this paper is divided into two parts and structured as follows. Part A describes the Work Productivity and Stress Questionnaire (WPSQ) development. Following the literature-based introduction of section 2, section 3 continues by describing the questionnaire construction, analysis approach, and data collection. In section 4, I describe the WPSQ factors following a factor analysis and show the validity and reliability. This section ends by comparing the WPSQ factors with the single-scale items of productivity. Part B, in turn, explores the recollection accuracy of the within-subject repeated measure using the WPSQ factors. In section 5, I discuss the relevant literature including recollection accuracy and memory to establish a theoretical framework. Section 6 continues by describing the multiple periods that the questionnaire targets and graphically display the hypotheses following the literature. This section also describes the analysis approach and models used to estimate the findings. Section 7 discusses the results, which include the descriptive trends, the within-sample differences and trends over time, and regression estimations of the observed recollection. Finally, section 8 concludes.

A. Constructing the Work Productivity and Stress Questionnaire

3.2 Background

The Health Work Questionnaire (HWQ) was developed aiming to uncover the relationship between health behavior and productivity (Shikiar et al., 2004). The HWQ measures workplace productivity, satisfaction, concentration, peer relations, and stress and enables the assessment of interventions at the workplace or health by measuring productivity on a multidimensional level. To date, the questionnaire has been used extensively in the literature and is considered a low-cost, valuable approximation of productivity in relation to worker health (Escorpizo et al., 2007; Healy et al., 2016; Hovinga et al., 2008; Lofland et al., 2004; Thorp et al., 2014).

Monitoring productivity using self-reported measures has become increasingly important during the shift from the office to working from home, especially as a result of the COVID-19 pandemic. For many employees, the pandemic-driven shift to the home environment was novel and abrupt. Employers, in turn, were unable to accurately and closely monitor work productivity and stress. Even basic communication was limited, and workers were often just sporadically seen through telecommunication. This lack of information increased the demand for periodical check-ups regarding productivity (Farooq & Sultana, 2021; Galanti et al., 2021; Zito et al., 2021). But straightforwardly asking how productive people feel might not reflect and cover the multitude of factors that comprise productivity. For instance, the physical environment itself could potentially hamper productivity beyond employee motivation. The extent to which the changed physical space, hardware, or coworker interaction quality affected productivity was unknown and required close monitoring (Davis et al., 2020; B. Wang et al., 2021; Yang et al., 2022). Moreover, the general physical and mental health as well as stress experienced by employees during this specific period (through work or nonwork factors) could potentially affect employees' work and nonwork satisfaction, and thereby productivity (Awada et al., 2021; Etheridge et al., 2020; Hallman et al., 2021). Therefore, productivity measures that include health-related factors may optimally capture the employee's sentiment in the move from the office to home (and back). The Health and Work Questionnaire covers both productivity and its relation to health (Halpern et al., 2001; Shikiar et al., 2001, 2004).

Although the Health and Work Questionnaire (HWQ) is well-cited and holds significant value as a tool for worker productivity and satisfaction assessment, scrutiny of the

HWQ uncovers three key limitations that request reconsideration. First, the HWQ was developed using a relatively small sample (N=294) with preselection criteria which might limit the external validity. Second, multiple studies comparing productivity and health measures critique the HWQ for neglecting explicit measures of absenteeism (Lohaus & Habermann, 2019; Mattke et al., 2007; Ospina et al., 2015). Absenteeism is defined as the unplanned absence from work due to illness or lack of concentration and is seen as an important factor in productivity loss (Johns, 2010). Finally, there is no information available on retest reliability or validity. How consistent and accurate the constructs are over time is disregarded, yet important for the (repeated) use in multiple domains (Beaton et al., 2009). Shikiar et al. (2004) underline that further questionnaire validation is needed before use as a primary measure. This paper sets out to mitigate these limitations, resulting in a better-identified, but still easy-to-apply worker satisfaction and productivity measure.

3.3 Methods

3.3.1 Questionnaire

The survey used in Shikiar et al. (2001) consists of three sections. First, all 30 items previously used in the original Health and Work Questionnaire construction process are included (Shikiar et al., 2004). These items are scored on a ten-point scale. The scales for each item are tailored to the questions, such that satisfaction items have a scale from "very dissatisfied" to "very satisfied", whereas others such as quantity or quality of work have a scale from "my worst ever", to "my best ever". These 30 items also include the six items that were excluded from the final version of the health and work satisfaction questionnaire proposed by Shikiar et al. (2004).

Second, the questionnaire includes multiple single-item scales. On a 10-point scale, ranging from "very" to "not all all", participants are asked to score "How productive are you?" and "How satisfied are you with work?", and "How suited is you job to do from home?". Additionally, participants indicate their willingness to continue to work from home ("yes" or "no"). Finally, participants are asked to rate, on a 7-point scale, ranging from "never" to "often", how often they take vacation days, call in sick, and take breaks during working hours. These latter items approximate "absenteeism."

Last, individual demographic information, as well as work and household factors, are measured, given that they may influence productivity adaption of working from home, or influence general productivity over time. It must be noted that this questionnaire is part of a larger study on the effect of indoor environmental conditions on working from home productivity (see chapter 4). Additional items that follow later in the questionnaire focus on the characteristics of the working from home office -- these items are independent and are thus not described or discussed in this paper.

3.3.2 Data collection

The data was collected during two periods, June 2020 and November 2020. In June 2020, 1,048 participants successfully completed the questionnaire. In November 2020, 772 of the 1048 participants from June 2020 completed the questionnaire for the second time. Thus, 276 participants from June 2020 did not participate in the second round. This attrition was compensated by collecting data from 230 new participants in November, bringing the total of the completed November 2020 surveys to 1,002. As shown in Table 3.1, for June and November 2020 combined, 772 repeated respondents and an additional 506 unique respondents (either only November or June 2020) led to 1279 participants being reached in total. Throughout this paper, we will refer to the first measurement in June as T1, and the second measurement in November as T2.

Table 3.1 Participant completion for both measures

Variables	Matched Scores	Unique Scores	Row Total
June 2020 (T1)	772	276	1,048
November 2020 (T2)	772	230	1,002
Both Measurements (Total)	772	506	1,278

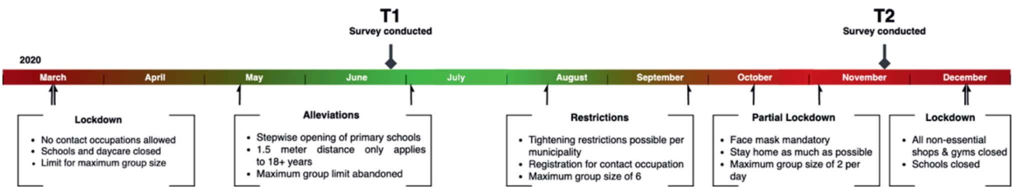
The 1,278 participants were approached via the Flycatcher panel. Flycatcher manages an academically-orientated, high-quality panel representing the cross-section of the Dutch population.[34] Flycatcher selected Dutch participants who met the inclusion criteria: office workers (with a minimum age of 18 years old), which worked at least part-time from home at the time of the survey. People without work, previously without work, or working exclusively

[34] See for example (Bults et al., 2011; Poelman et al., 2021; Smeets et al., 2015)) for well published studies using the Flycatcher panel.

from the office are excluded from the sample. Participation is reimbursed. The research setup was reviewed and approved by Maastricht University's Ethical Review Committee Inner City Faculties (ERCIC_195_09_06_2020).

The COVID-19 restrictions that applied during this time determined the majority of homework in our sample. Figure 3.1 shows an overview of the development of restrictions in the Netherlands from March 2020 (the start of the restrictions) until December 2020. In this Figure, measurements T1 and T2 are highlighted with respect to COVID restrictions. It is important to note that, although the data in June (T1) was collected during a period of alleviation from the first lockdown restrictions, working from home was still the default at that time. There is no reason to believe that self-selection in working from home applies to this measurement. In November (T2), all previous easements (i.e. alleviations) were reversed whilst moving into the second lockdown.

Figure 3.1 Timeline of COVID-19 restrictions with respect to the data collection



3.3.3 Analysis

The analysis of section A follows three steps. First, I reconstruct and assess the performance of the original HWQ, I recreate the factors of the HWQ based on Shikiar et al. (2004). I asses the internal consistency (using Cronbach's alpha) and the scalability (using Loevinger's Hj coefficient) of these factors for both the T1 and T2 measurements.

Second, the items of the original survey are subjected to principal component analysis which results in a revised questionnaire. This principal component analysis of the items, without recreating the original factors of the HWQ, identifies the optimal number of factors that explain the most variation between all the items. After identification of the optimal number of factors, the items are fitted to one of these finite factors, using orthogonal varimax rotation (Merenda, 1997).

Third, I validate the revised Work Productivity and Stress Questionnaire and asses its internal consistency. The item allocation fit is evaluated using both metrics and content (external) validity. I assess the metrical fit of the factors using the internal consistency of the factor (using Cronbach's alpha), the scalability (using Loevinger's Hj coefficient), and the individual inter-item correlations. The Loevinger's Hj coefficient, specifically, estimates the accuracy for the ability of items to order the respondents in the measured latent trait (Roskam et al., 1986). The external validity, of face value, assesses the theoretical belonginess these items have in the group of items in that factor. This also leads to the categorization of the factors, and consequent factor labelling. Reliability over time is assessed by correlating the T1 and T2 measurements of each factor. Additionally, congruent and discriminatory validity is investigated by correlating two specific factors (productivity and stress and irritability) with alternative (both single-scale and validated) measures.

3.4 Results

3.4.1 HWQ Validation

Although the original HWQ tested 30 items, only 27 items were included in the original HWQ questionnaire after validation. First, I include these 27 items and validate the original HWQ questionnaire by recreating the HWQ factors a priori. The factors are formed by allocating each item to its factor as described in the paper. Table 3.2 shows the descriptive scoring for the constructed HWQ factors for both T1 and T2. Note that the "Supervisor Relations" factor is indicating missing values since some workers in our sample do not work with or for a supervisor (for instance, in case they are their own boss or head of a company).

Table 3.2 Descriptive Statistics HWQ scales in the sample

	N	Mean	SD	Min	Max	alpha	Loevinger H	Loevinger Hj-min
Panel A. T1								
Productivity	1,048	7.32	1.02	2	10	.90	.47	.27
Concentration/Focus	1,048	7.02	1.76	1	10	.81	.53	.47
Supervisor Relations	965	6.98	1.61	1	10	.83	.74	.74
Nonwork Satisfaction	1,048	7.10	1.10	3	10	.61	.35	.31
Work Satisfaction	1,048	6.84	1.07	3	10	.64	.32	.27
Impatience /Irritability	1,048	3.65	1.71	1	10	.80	.63	.60
Panel B. T2								
Productivity	1,002	7.19	1.18	1	10	.92	.53	.35
Concentration/Focus	1,002	6.85	1.97	1	10	.81	.54	.48
Supervisor Relations	915	6.86	1.72	1	10	.81	.70	.70
Nonwork Satisfaction	1,002	6.43	1.42	1	10	.62	.36	.34
Work Satisfaction	1,002	6.47	1.28	1	10	.64	.32	.28
Impatience /Irritability	1,002	3.57	1.89	1	10	.82	.67	.63

A first look at the descriptive statistics shows productivity scores well above 7, on average, for both measurement periods (mean = 7.19, SD= 1.18; mean = 7.32, SD = 1.02, respectively). Productivity measured by Barone Gibbs et al. (2021) using the same questionnaire, shows a comparable mean and standard deviation (mean = 7.1, SD=1.3) for homeworkers during the pandemic. They find that this score significantly decreased compared to pre-pandemic (mean = 7.6, SD=1.1; p<.001). Although Barone Gibbs et al. (2021) find no difference in the impatience (and irritability) scale before and during the pandemic, their score during the pandemic is noticeably lower compared to this sample (mean = 1.8, SD = 1.8). Note that the items in the impatience (and irritability) scale are scored such that a low score indicates low levels of impatience or irritability. A high productivity score is in turn, unsurprisingly, accompanied by a low impatience score.

To what extent the questionnaire accurately captures the health and work constructs, is indicated by the metrics on the right-hand side of Table 3.2. The internal consistency of the subsequent scales is expressed by Cronbach's alpha. On a scale of 0-1, the alpha indicates how closely related a set of items are as a group, with 1 indicating perfect relation, and 0 completely unrelated. Table 3.2 shows that work and nonwork satisfaction score poorly, with alphas below the generally accepted threshold of 0.70 (De Vet et al., 2011). During the original construction, Shikiar et al. (2004) reported the impatience (and irritability) factor to be of relatively low alpha at 0.72. However, this sample does not confirm this finding.

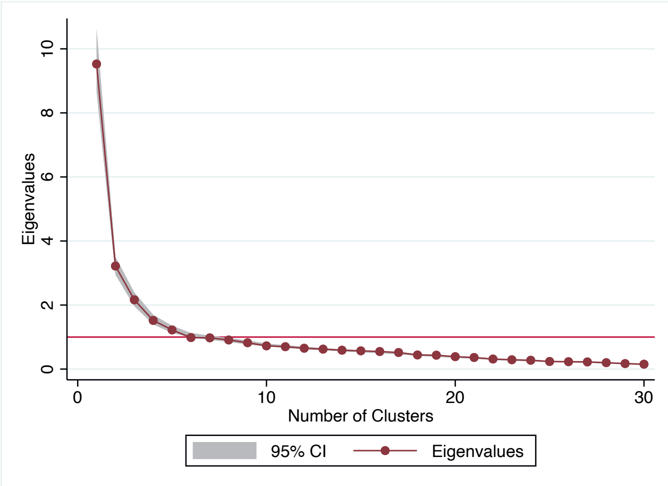
Next, the scalability of the factor, the degree to which individual items deviate from the collective scoring of items, is assessed using Loevinger's Hj coefficients (Schäfer et al., 2021). This metric can be calculated for the full scale (H; the degree to which a factor fits the questionnaire's construct) as well as for each individual item (Hj; the degree to which an item fit within a factor). Generally, a factor score $0.3 < H < 0.4$ is considered a poor factor, and an item score Hj is that $0.3 \leq H$ is considered a poor-fitting item for that factor (Stochl et al., 2012). In both panel A and B, work and non-work satisfaction show a poor scalability score, with work satisfaction having an item with Hj below 0.3 for both scales. In addition, for T1, productivity holds an item that fits relatively poor in the overall scale, although the scale H remains above 0.4.

Overall, both panels show that productivity has a high internal consistency, but the factors might contain at least one item that harms the reliability. Moreover, both non-work as well as work satisfaction factors do not meet the minimal requirements on both internal consistency or scalability. Newly constructed factors based on the existing items by means of principal component analysis thus have a high potential to increase the validity of productivity assessment tool.

3.4.2 Principal Component Analysis

A principal component analysis was conducted on all 30 items of the questionnaire. Figure 3.2 shows five factors with an eigenvalue above the threshold line of 1 (1.22 for the fifth factor and 0.98 for sixth factors). These five factors explain 59% of all variation (for the the eigenvalues, see Appendix Table 3.12). This judgment, in line with the Kaiser Criterion (components with eigenvalues over 1 are meaningful), coincides with the scree plot analysis (Yeomans & Golder, 1982). The rule-of-thumb states that the 'elbow' joint, where the scree plot shallows out of an angle, indicates a decreased value of additional components (Cattell, 1966). After component 5, limited additional variance is explained by component 6 onwards. Identification and further analysis of the five-factor solution of the internal consistency of the factors and the scalability of the individual item further confirm this cutoff.

Figure 3.2 PCA scree plot of Eigenvalues



An orthogonal varimax rotation displays the optimal allocation of each item with a respective factor. Table 3.3 shows factor loadings and item allocation per factor. Stevens (2012) suggests that critical values indicating relevant loadings depend on the sample size. Based on a sample of 1,000 participants, factor loadings below 0.162 are not considered significantly relevant, using an alpha level of 0.01 (two-tailed). Therefore, values below 0.162 are colored light grey in Table 3.3.

The scalability assessment indicates one outlying item that showed Loevinger Hj-min below 0.3: WPSQ item 29 was dropped from its respective factor. Note that two items that were originally dropped now remain in the framework (WPSQ items 28 and 30; for a complete overview of the WPSQ versus HWQ item and factor allocation, see Appendix Table 3.13). One item ("*What is your view of your quality of work for this week*"; WPSQ ID 4) loads comparable on both factor 1 and factor 2. Since deleting this item from the scale will harm the internal consistency (Cronbach's alpha decreases from 0.88 to 0.87), I categorize this item in factor 1, as it fits the overall construct as well as the other items in that scale increasing the face and content validity.[35]

[35] Performing an identical factoring for other measurement points in our data (not reported) shows roughly random dominate allocation based on factor loading to either the first or the second factor. Overruling the factor loading based allocation based on construct validity is not uncommon. The original by Shikiar et al. (2004) has a similar conflict, although they do not mention their consideration (see the original paper Table 1, HWQ item 12B)

Table 3.3 Factor analysis of the Work Satisfaction and Productivity Questionnaire (WPSQ)

WPSQ ID	Description	Factor Loadings				
		1	2	3	4	5
		Productivity	**Productivity by others**	**Stress and Irritability**	**Peer Relations**	**Nonwork Satisfaction**
1	Your view of your efficiency this week	0.30	0.16	0.08	-0.09	-0.01
4	Your view of your quality this week	0.19	0.23	0.02	-0.08	0.02
7	Your amount of work this week	0.23	0.22	0.11	-0.15	0.03
10	Rate highest level of efficiency	0.28	0.12	0.03	-0.12	0.02
11	Rate lowest level of efficiency	0.29	0.03	0.15	-0.02	-0.07
13	Frequency of boredom at work	0.31	-0.07	-0.11	0.03	-0.08
14	Frequency of low concentration at work	0.34	-0.06	-0.14	-0.09	-0.01
19	How satisfied with your job?	0.32	-0.05	0.02	0.21	-0.03
21	How satisfied with work environment?	0.26	-0.08	0.08	0.09	0.03
22	How rewarding was work in past week?	0.36	-0.06	0.05	0.18	-0.06
2	Supervisor's view of your efficiency	-0.02	0.32	-0.02	0.21	-0.05
3	Coworkers' view of your efficiency	0.01	0.34	-0.02	0.02	0.04
5	Supervisor's view of your quality	-0.07	0.38	-0.07	0.17	-0.05
6	Coworkers' view of your quality	-0.06	0.40	-0.07	0.00	0.02
8	Supervisor's view of your amount of work	0.04	0.36	0.05	0.02	-0.01
9	Coworkers' view of your amount of work	0.00	0.40	0.02	-0.12	0.05
12	Frequency of restlessness at work	-0.23	0.10	0.23	0.08	-0.08
15	Frequency of being too exhausted to do work	-0.16	0.02	0.28	0.09	-0.03
25	Frequency of annoyance with coworkers	-0.04	0.03	0.41	-0.10	0.05
26	Frequency of impatience with coworkers	0.05	-0.01	0.48	-0.06	0.02
27	Frequency of conflicts with coworkers	0.07	-0.08	0.44	-0.02	0.05
28	How stressed are you?[2]	-0.10	0.04	0.27	0.12	-0.23
30	Frequency of unable to complete tasks[2]	-0.08	-0.05	0.26	0.02	0.07
16	How satisfied with supervisor relationship?	0.01	0.02	-0.01	0.57	0.05
17	How easy to communicate with supervisor?	-0.01	0.01	-0.07	0.52	0.00
24	How satisfied with coworker relationship?	0.08	0.00	0.04	0.27	0.22
29	Experienced control on work[1,2]	0.13	-0.07	0.19	0.21	-0.07
18	How rewarding was personal life?	0.11	-0.04	0.00	-0.04	0.44
20	How easy to communicate with family/friends?	-0.03	0.02	-0.02	0.03	0.54
23	How satisfied with relationships?	-0.05	0.01	0.04	0.04	0.60

[1]This item is excluded from the WPSQ scale.
[2]These items are excluded from the original HWQ scale. Loadings below 0.162 are not considered significant and are therefore noted in light grey.

3.4.3 WPSQ Item Analysis and Factor Description

The item allocation fit to each factor is assessed from both a metric and theoretical point of view (Rattray et al., 2004). Two metrics assess the item fit: the internal consistency of the factor, and the individual inter-item correlation. For internal consistency, no item's internal consistency falls below the critical Cronbach's alpha of 0.5 for each factor (Kline, 1993).[36] This indicates that the items fit well within their respective factor from a technical perspective. Inter-item correlation provides information on whether an individual item covers sufficiently unique variance within their factor. If this correlation between all items in the factor is too high, the items could be considered redundant. If the correlations are too low, they are no longer homogenous enough to address the same construct. Ferketich (1991) suggests an optimal correlation between $0.3 \leq \alpha \leq 0.7$. For all factors, only two items have an inter-item correlation larger than 0.7 (specifically, 0.8): WPSQ 19 and WPSQ 22 (for all inter-item correlation matrices per factor, see Appendix Table 3.14). Further inspection shows that these two items indeed have conceptual overlap ("*How satisfied are you with your job?"* and "*How rewarding was your work in the past week?"),* yet differ crucially in the timespan they cover. Job satisfaction focuses on general satisfaction, whereas the level of reward experienced in the past week is specific to this period. Although it is not surprising that these items correlate in our sample, it cannot be assumed that this is generally true outside our context. As such, these items are both retained in the factor.

For theoretical assessment, clarity and relevance are the key determinants of the item fit (Rattray & Jones, 2007). Based on the content of the relevant items, the factors are labeled productivity, productivity by others (productivity as seen by others), stress and irritability, peer relations, and nonwork satisfaction. Productivity includes items that focus on efficiency, quality, and quantity of work. It also includes items that cover the rewarding and satisfying nature of the work. Finally, it contains items that assess the frequency of boredom and low concentration at work. Note that the last items are scored reversely on this scale, meaning that low boredom contributes to a high productivity score. Productivity by others covers the perceived or predicted score on efficiency, quality, and quantity by both supervisors and coworkers. The category Stress and Irritability includes questions such as *'How stressed are you?'*, but also items describing conflicts and annoyance with coworkers. These negatively scored items differ from boredom and low concentration to the extent that the latter can be correlated with motivation and work satisfaction, whereas the former items pertain more to exhaustion-related stress and consequent irritability. The scores of these items are not reversed, such that high scores indicate high levels of stress and irritability. Peer relations does not include negative conflicts with coworkers, but focuses on the ease of communication and satisfaction of coworker and supervisor relations. Finally, non-work satisfaction covers items such as '*How rewarding is your personal life?*', '*How easy*

[36] Table is not reported in this chapter

is it to communicate with your friends and family?', and '*How satisfied are you with your relationships?*'. In line with the description, all items are considered relevant additions to their respective factors.

3.4.4 WPSQ Factor Reliability

The reliability of the factors is determined by the internal consistency of the factors (Kline, 1993) and the stability of the measure over time (Johnson, 2001). The internal consistencies of all the WPSQ factors are shown in Table 3.4. For both measurements, and each factor, the internal consistency Cronbach's alpha does not fall below 0.73. This is an improvement as compared to the original HWQ scale in Table 3.2, where factors often dropped below 0.65. Similarly, the stability of the factors, indicated by Loevinger's H, falls only once marginally below 0.4. Compared to the HWQ, the problematic factors Work and Nonwork Satisfaction no longer decrease the overall reliability of the survey.

Table 3.4 Descriptive Statistics WPSQ scales in the sample

	N	Mean	Std. Dev.	Min	Max	alpha	Loevinger H	Loevinger Hj-min
Panel A. T2								
Productivity	1002	6.84	1.28	1	10	.88	.45	.31
Productivity by others	957	7.55	1.25	1	10	.90	.62	.60
Stress and Irritability	1002	3.82	1.63	1	9	.83	.43	.36
Peer Relations	968	6.65	1.59	1	10	.77	.54	.46
Nonwork Satisfaction	1002	5.99	1.59	1	10	.75	.51	.45
Panel B. T1								
Productivity	1048	6.94	1.07	2	10	.85	.38	.22
Productivity by others	1001	7.65	1.15	1	10	.89	.62	.59
Stress and Irritability	1048	3.82	1.46	1	9	.81	.40	.30
Peer Relations	1012	6.89	1.41	1	10	.78	.57	.47
Nonwork Satisfaction	1048	6.98	1.17	2	10	.73	.49	.42

Note that for panels B and C, Productivity has one item that has a Loevinger Hj below .3. The same item also has the lowest Hj for Productivity in Panel A. Deleting this item, WPSQID 21 ("How satisfied with work environment?"), would not improve the internal consistency of the scale measured in Panel A, and only increase the productivity scale in Panel B and Panel C with .01 for both factors.

To assess the reliability over time, I compare the scores of the factors in the November measurement (T2) with the scores of the same factors in the June measurement (T1). Appendix Table 3.15 shows the correlation matrix between the measurement in November and June, for all factors. The correlation between the T1 measurement and the T2 measurement are moderate, between 0.42 and 0.59 (average of 0.53; see Appendix Table 3.15). This correlation is not particularly strong, but it must be noted that these two metrics are collected with a significant amount of time in between, during a period in which productivity was subject to adjustment. These measurements are therefore expected to change over this time period, as the experience of people with working from home grew. It is expected that, after the adaptation phase of working from home, the correlation will increase even more.[37]

3.4.5 WPSQ Factor Validity

In order to assess how valid the items and factors are in testing the underlying construct, I consider face and construct (congruent and discriminatory) validity.[38] Face validity pertains to the level of clarity of which a questionnaire and its items appear to relate to the targeted construct. In other words, the items should not only measure the construct, but also appear to measure the constructs (Mosier, 1947). Table 3.3 shows all the individual items per factor, and as previously discussed in the factor description, all items related naturally to the overarching constructs.

The construct validity describes the correlation with metrics that measure the same underlying construct (congruent) and the lack of correlation with metrics that measure unrelated constructs. For the congruent validity, I compare the productivity factor with direct measures of productivity and work satisfaction. As shown in Table 3.5, the WPSQ productivity factor strongly correlates with the general self-reported level of productivity, and the self-reported level of work satisfaction (Cronbach's alpha of 0.73 and 0.79, respectively). Moreover, the WPSQ factor correlation with both the single question metrics consistently exceeds that of the HWQ productivity factor.

Table 3.5 Correlations of Productivity

Variables	Productivity (WPSQ)	Productivity (HWQ)	How Productive at work	How satisfied with work
Productivity (WPSQ)	1			
Productivity (HWQ)	0.81	1		
How Productive at work	0.73	0.66	1	
How Satisfied with work	0.79	0.54	0.56	1

Comparatively, Table 3.6 shows that the WPSQ factor 'Stress and Irritability' is strongly correlated with burnout propensity (Cronbach's alpha of 0.71). Again, the WPSQ is

[37] Later, in part B, I will discuss the value of consistency over time in self-report use of this metric.
[38] Ideally, predictive validity, the extent to which the factor predicts the objective behavior, would be included as well. Unfortunately, the nature of the survey limited the data to self-report, which prohibited me from comparing the scores to, for instance, realized productivity.

consistently higher correlated with burnout propensity than the HWQ 'Irritability' factor. I also add break time, days off of work, and sick days. They are often associated with stress and burnout propensity and are particularly valuable as the HWQ is often criticized for not including explicit measures of absenteeism (Mattke et al., 2007). The comparison in Table 3.6 indicates that the HWQ factor 'Irritability' poorly correlates with the separate direct measures of absenteeism. With the exception of breaktime, the WPSQ factor 'stress and irritability' correlation with these indicators is not only higher, but also closely mimics the correlation between these indicators and burnout propensity.

Table 3.6 Correlations of Stress and Burnout

Variables	Stress (WPSQ)	Irritability (HWQ)	Burnout Propensity (MBI)	Breaktime	Days off work	Sickdays
Stress (WPSQ)	1					
Irritability (HWQ)	0.84	1				
Burnout Propensity (MBI)	0.71	0.51	1			
Breaktime	-0.08	-0.08	-0.10	1		
Days off work	-0.11	-0.07	-0.14	0.30	1	
Sickdays	0.21	0.11	0.28	0.07	0.13	1

Comparing stress and irritability, burnout propensity, and productivity also provides insights into the discriminant validity. Ideally, productivity and stress should be negatively correlated (Halkos & Bousinakis, 2010; Salehi et al., 2010).[39] Indeed, we find the strongest negative correlation between the two WPSQ factors Productivity and Stress (correlation of -0.55; see Appendix Table 3.16). This indicates that, although the factors are highly correlated with similar constructs, the discriminatory value is highest for the factor, making it most distinctive of all our metrics. Comparing all factors with each other, and across multiple times, only weak correlations are detected (Cronbach's alpha below 0.40), with only incidental exceptions (see Appendix Table 3.16).

In conclusion, the WPSQ factors show a clear improvement over the original HWQ factors. The new PCA resulting allocation of items is relevant and fitting. The factors have a high face and construct validity and display improved internal consistency per factor indicating psychometric reliability over the original HWQ factors.

[39] Some research suggests an inverted U-shape between stress and productivity, implying that some degree of stress increases productivity (e.g. Salehi et al. 2010). As this is beyond the scope of this paper, I treat the relation between stress and productivity classically linear.

3.4.6 WPSQ versus Single-item Scale

The comparison between WPSQ's productivity factor and the single-item scales ("How productive/satisfied are you at work?") also provides insight into the value of these single-items as approximations of productivity. The single-item scales suffer from the lack of variance that would enable them to cover the multitude of factors influencing productivity. Moreover, whereas the single question is somewhat sensitive to complacent responses, the multi-item constructs seem more likely to capture nuances within people. Generally, multi-item scales are preferred over single-item scales to avoid misinterpretation, biased responses, and reduces measurement error (Bowling, 2014; Grove et al., 2012). Interestingly, Table 3.7 (a selection of variables from Appendix Table 3.16) shows an example of the added value of the WPSQ factors over the single-item estimators. The divergent validity of the WPSQ productivity metric was previously discussed by correlating it with stress, burnout propensity, as well as other indicators. When substituting the WPSQ productivity measure with the single-item scales, as can be seen in Table 3.7, the results show that both have comparable correlations. High productivity is correlated with lower stress, lower burnout propensity, fewer sick days, and longer break times. However, for all these expected correlations, the WPSQ factor shows a stronger and more stable trend. Specifically, the single-item productivity scale seems to resemble the WPSQ factor when it comes to break time and sick days, whereas the single-item work satisfaction scale seems to resemble the WPSQ factor's correlation pattern with stress and burnout propensity. As there is no strong a priori rationale for the interaction between work satisfaction and productivity with these two specific sets of indicators, it can be assumed that the WPSQ provides a more integrative approximation of all related concepts.

Table 3.7 Divergent Correlations of WPSQ versus Single-item Scales

Variables	Productivity (WPSQ)	How Productive at work	How Satisfied with work
Stress (WPSQ)	-0.55	-0.37	-0.48
Burnout Propensity (MBI)	-0.50	-0.31	-0.46
Breaktime	-0.11	-0.13	-0.03
Days off work	0.00	-0.06	0.03
Sickdays	-0.21	-0.17	-0.13

Nevertheless, the benefits of single-item measures are clear: they are simple to implement and ask little time to complete. The results in this paper indicate that the predictive value of single-item scales might be lower, yet largely comparable to their more extensive WPSQ counterpart. When brevity is wanted or needed, the single-item scales can be considered as an alternative. Moreover, the previously discussed distinct patterns of work satisfaction and productivity with the specific indicators suggest that they both inform about different factors. However, I cannot conclude that each single-item scale can approximate different factors based on these results. Further research needs to formulate and confirm clear a priori expectations about the relation between the single-item scale and the targeted factors to justify the single-item scale isolated use. Until then, I recommended using both single-item scales (work productivity and work satisfaction) together.

B. Using the Work Productivity and Stress Questionnaire Retrospectively: Recall bias

To illustrate how the Work Productivity and Stress Questionnaire can be deployed, I measure the factor's trends over time during the working-from-home shift in our Dutch sample described in Part A of this chapter. The shift coincided with a volatile period where many changes in office policies were forcefully implemented without proper piloting opportunities. More than ever, productivity and work satisfaction monitoring was used to ensuring employee happiness, wellbeing, and retention during this turbulent time (Farooq & Sultana, 2021; Galanti et al., 2021; Zito et al., 2021). The dependency on self-reported productivity, however, requires accuracy in reporting and recollection. Recollection accuracy is important even if recollection is not explicitly targeted since it functions as a reference point for comparative scoring. For instance, in order to estimate the current working-from-home productivity, you first recollect your general productivity (likely at the office). Yet, whether recollection of productivity over a rapidly changing period is indeed accurate, is questionable. Many strains of literature question the accuracy of autobiographical memory and the influential role of memory biases (for an overview, see Schacter, 1999, 2022). One key observation is that the current state-of-mind tends to paints the past (Bower, 1992). However, the productivity domain often falls outside of the scope of the literature. This is not surprising, as until the recent forced shift to home, objective or second-party reports on productivity were readily available. Since the dependency on self-report is now steadily increasing, I investigate whether recollecting productivity (and working-from-home factors) is accurate over time, or whether existing recall bias influences the way we evaluate working-from-home.

3.5 Recall Bias Literature

The pandemic forcefully submerged all workers in the working from home experience. This meant working-from-home evaluations were for the first time independent of individual preferences or predispositions with regard to homeworking. On the other hand, using the pandemic as a shock to evaluate work from home comes with limitations. Specifically, self-report, essential during this shift, could be less trustworthy than desired during a rapidly changing period. For example, the potential negative impact of the pandemic on one's personal life might spill over to productivity, causing one to be or feel unproductive (Collins, 2020). Moreover, retrospection over highly volatility periods generally decreases recollection accuracy (Arslan et al., 2021; Cross et al., 2021; Ogden, 2021; Puente-Díaz & Cavazos-Arroyo, 2022) [40]. For instance, valuing the time spend at home with family increases the perceived working from home satisfaction (Barrero et al., 2021; Brunelle & Fortin, 2021; Ipsen et al., 2021). This could negatively carry-over to the recollection of the retrospective work satisfaction at the office before the pandemic. Without identifying these errors in recollection, or recall bias, in the domain of productivity and working from home, the errors might incorrectly influence how we evaluate working-from-home.

Generally, a recollection inaccuracy (or recall bias) is an often-observed phenomenon in autobiographical memory in cohort studies (Coughlin, 1990). For example, people commonly experience personal growth over time. This growth is often artificially achieved by adjusting the recollection of the past performance downwards, although actual performance might have been equal or better (Wilson & Ross, 2000). On average, the net effect of a recall bias is the exaggeration of the magnitude of difference (Raphael, 1987). In health research, treatment groups often recall a period prior to treatment as worse compared to a control group, enlarging the effect of the treatment (Holmberg et al., 1996). Likewise, recalling the shift from the office to working from home could enhance the recalled change in productivity. However, in contrast to health research, the universal shift to home due to the pandemic eliminated any non-treated control groups needed to detect or test for recall bias. Going unnoticed, a recall bias could lead to misidentifying a memory error as a significant change on, for instance, productivity.

[40] Note that I distinguish between retrospection and recollection such that retrospection is the act of looking back, whereas recollection is the recalling using memory. Since there is not always a score from a retrospective period (before the pandemic, for instance), I implicitly require participant to use recollection to recall their retrospective score.

The implicit theory framework explains why a recall bias could also exist in the productivity context. According to this framework, people form beliefs about their own attributes which, in turn, could shape or influence memory (Ross, 1989). Recalling attributes in the past, such as productivity, starts from the most accurate and recent reference point: the current situation. To get to past performance, people generally take their current performance and adjust according to their beliefs about the stability of this attribute over the past time. Implicit theory suggests that, if people believe this attribute to be relatively stable, they are likely to judge the past period as stable too, and will score their performance closer to their current performance (Conway & Ross, 1984). If people expect that something (for instance, a pandemic) drastically changed performance, they will revise their score to exaggerate the difference. The need to be consistent drives the general adjustment. Note that this is reversed causality: the scores will be adjusted so that they reflect the expected change, instead of concluding change based on different scores.

Whether recall bias will materialize is not guaranteed and depends on multiple factors. For instance, the sensitivity generally increases with the length of the recalled period (i.e. longer periods are more sensitive) and the intensity of a past period (i.e. more intense shifts are more sensitive; Bryant et al., 1989; Coughlin, 1990; Weinstock et al., 1991; Wilson & Ross, 2003). Moreover, accuracy differences resulting from recall bias do not always influence study results (Drews & Greeland, 1990). It is, therefore, not certain that work-related factors such as productivity are actually subject to recall bias in recollection.

To fill the gap in the literature, and understand how sensitive productivity is to recall bias, I explore productivity recollection during a prolonged work-from-home period. Comparing recollected scores with the past scores from the same period enables a first indication of whether recollected, self-reported productivity is accurate, or subject to recall bias. In the following section, I investigate the development of Work Productivity and Stress Questionnaire factors over time following the start of the pandemic. In doing so, the comparison of productivity and stress can be compared over time, including pre-pandemic (at the office) and during pandemic (at home), that uncovers first insights into the success of working from home in this context. I additionally explore the recollection trend of both measurements, which for a large part overlap. This overlap enables me to infer recall bias when comparing recollection scores with the target scores, but also shows the recollective trends of both measurements.

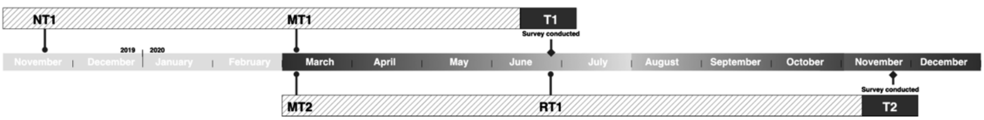
3.6 Methods

3.6.1 Questionnaire

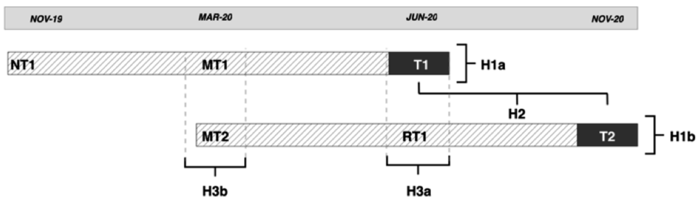
In order to monitor the productivity and work satisfaction during the transition from the office to work, I ask participants during both measurements to report on Work Productivity and Stress Questionnaire factors at the time of the measurements (June and November 2020; T1 and T2 respectively), but also to recollect the scores on the factors for two previous periods. Figure 3.3A schematically shows the two measurement points T1 and T2, and their retrospective periods. In June 2020, participants scored all questions for their current state (June 2020: **T1**), for the beginning of the pandemic in March (**M**arch scored at **T1**: **MT1**), and before the pandemic (**N**ovember 2019 period **T1**: **NT1**). In November 2020, the same participants were asked to score their current status (November 2020: **T2**), their status during June 2020 (**R**ecollection of **T1**: **RT1**), and again, during the beginning of the pandemic in March (**M**arch period score at **T2**: **MT2**). This structure enables a comparison between a direct measure (T1) with a retrospective measure (RT1) and two retrospective measures for the same period (MT1 versus MT2).[41]

Figure 3.3 Schematic Overview

A. Measurement periods



B. Schematic overview hypotheses



Note. Figure 3.3A shows the measurement period as well as the retrospective period for each measurement, T1 and T2. The month bar is colored based on the severity of the lockdown measures in the Netherlands, such that darker indicates stricter rules. Figure 3.3B schematically displays the relationship between each hypothesis and the targeted scores collected during either T1 or T2. Note that T1 and T2 are identical to Figure 3.1.

[41] Note that I assume that other questions included in the survey, used for Chapter 4, are not interfering with the recollection. It is reasonable to assume that the indoor environment characteristics are not likely to change significantly over time, nor are likely to influence recollection. Additionally, pre-pandemic (NT1) could be compared to current (pandemic; T1). This comparison is discussed in Chapter 4 and is not included in this paper. As such, NT1 will not be discussed further.

3.6.2 Hypothesis and Analysis

In this exploration, I test the self-reported trend over time for each separate measurement, the within-sample development between the two current measurements (T1 and T2), and the recollection accuracy. Figure 3.3B shows an overview of these three main hypotheses. Hypothesis 1a and 1b first examine whether the retrospective scores of all WPSQ factors differ significantly from their respective current state. This hypothesis explores whether significant changes have been reported between each reported time period (for instance, productivity in March 2020 was different from the productivity in November 2020). If WFH did not have an influence on self-reported productivity in the sample, all three measurements in T1 and T2 are expected to be equal. This is tested by means of a non-parametric repeated-measure within-subject ANOVA alternative, the Friedman Test. A significant difference, implying that not all scores are equal, will be further examined in a post-hoc analysis using the extended Mantel-Haenszel (Cochran-Mantel-Haenszel) stratified test of association.

Hypothesis 2 determines whether the current state of all factors in T2 changed with respect to their respective state in T1, using a within-subject non-parametric Wilcoxon signed rank test. This hypothesis tests from a classic repeated-measure consistency perspective. In contrast to hypotheses 1a and 1b, hypothesis 2 does not measure stability in retrospective scores for the same measurement, but rather compares stability between the two separate measurement periods.[42] Again, if the switch to WFH would not have had an effect on self-reported measures, T1 and T2 would be equal.

Finally, the recall bias is examined in two steps. First, a within-subject non-parametric Wilcoxon signed rank test will determine whether a bias exists, by testing whether RT1 significantly differs from T1. Hypothesis H3a identifies a recall bias, since I compare the recollection of T1 in the T2 measure (**R**ecollection **T1**: **RT1**), with the actual scores at that time: T1. Comparing two recollection measures MT1 and MT2, without an objective 'current' score, shows the perseverance of the bias at H3a. Note that for both hypotheses, the effect of WFH is no longer relevant, but I estimate the consistency in recollection. If people accurately recollect their scores in the last, both H3a and H3b should be equal. Second, I attempt to identify the root of the recall bias of H3a. I use multiple linear regressions (MLR) to assess the relationship between the retrospective score (RT1) and both the target score (T1) and the current score (T2) of all respective WPSQ factors, using the following models:

[42] Note that hypotheses 1 and 2 do not look at recollection accuracy.

$$RT1_i = \alpha_0 + \alpha_1 T1_i + \alpha_2 T2_i + \alpha_3 T1_i T2_i + \alpha_4 Controls_i + \varepsilon_i \quad (1)$$

For each WPSQ factor, model 1 estimates to what extend each respective WPSQ factor *j* is predicted by the score of that same WPSQ factor at T1 and T2. An interaction term is added, acknowledging the potential relationship between T1 and T2. This interaction term controls for the influence of one predictor on the other. Introducing this in the model, will unveil the average effect of each predictor (T1 or T2) whilst keeping their relationship constant. Additional work, household, and individual controls are added to control for individual or contextual factors that might interfere with the respective WPSQ scores or memory accuracy. This included gender, age, income bracket, education bracket, children (at home during office hours), deadlines and experienced control by work, work suitable for WFH, and willingness to continue with work from home.

$$\Delta(T1 - RT1)_i = \beta_0 + \beta_1 T2_i + \beta_2 Controls_i + \varepsilon_i \quad (2)$$

$$|\Delta(T1 - RT1)_i| = \gamma_0 + \gamma_1 T1_i + \gamma_2 T2_i + \gamma_3 T1_i T2_i + \gamma_4 Controls_i + \varepsilon_i \quad (3)$$

A second model (2) aims to explain the relative bias magnitude between RT1 and T1. In this model, the delta between the target score (T1) and the recollected score (RT1) is predicted by the supposedly independent score T2. Since a coefficient from this model is relative to the under- or overestimation as well as the increase of decreased score of T2 from T1, model 3 finally explains the absolute delta between the target score (T1) and the recollected score (RT1). The absolute delta shows the net inaccuracy: the higher the absolute delta coefficient, the higher the difference between T1 and RT1 scores.

Additionally, model 4 explores the underlying effect of the current measures on any discrepancy between the retrospective scores from both measurement (MT1 versus MT2). Comparable to model 3, model 4 estimates the effect of the absolute delta between the current scores on the absolute delta between both retrospective scores:

$$|\Delta(\mathrm{M}T1 - MT2)_i| = \delta_0 + \delta_1 |\Delta(T1 - T2)_i| + \delta_2 Controls_i + \varepsilon_i \quad (4)$$

Finally, I explore the discrepancy in retrospective scores of both measures. The WPSQ scores for March collected at both T1 and T2 are on retrospection. In order to explore the recall bias for these scores, I estimate the predictive value of the absolute delta of both current scores on T1 and T2 on the absolute delta of both retrospective scores of March (MT1 and MT2, respectively).

For all non-parametric paired comparison tests, I apply the most conservative multiple testing procedure (e.g., Bonferroni correction; Thissen et al., 2002). The odds of finding a false significant result when testing 5 simultaneously factors at a 5% significance α level can be calculated using the binominal formula:

$$P(false\ positive) = 1 - (1 - 0.05)^5 = .23 \tag{5}$$

Without multiple testing correction, the formula shows that the odds of obtaining a false positive equal roughly 23%. The Bonferroni correction resets the significant level such that, for multiple tests, the odds of obtaining a false positive return to a 5% level. By dividing the α by the number of multiple tests (5 factors, in this case), the renewed p-value threshold equals 0.01:

$$P(BF\ correction\ false\ positive) = 1 - (1 - 0.01)^5 = .0495 \tag{6}$$

Throughout the paper, paired tests apply the following threshold conversion: α=5% equals 0.01 (indicated by one star), α=1% equals 0.002 (indicated by two stars), and α=0.1% equals 0.0004 (indicated by three stars).

3.7 Results

3.7.1 WPSQ factor trends over time

Plotted in Figure 3.4, the average scores of all participants show that almost all scores dropped immediately after the start of the pandemic in March, and, although improving over time, do not completely recover at the latest measurement (T2). Although Figure 3.4 plots the mean trends and suggests a lack of significant difference between time periods, hypothesis 1 focuses on the within-subject variation. The average scores of all WPSQ factors for all collected time periods are shown in Table 3.8. A non-parametric repeated-measure within-subject

ANOVA alternative, Friedman Test, shows that all scores within each measurement differ significantly over time. Conclusively, this implies that T1=MT1=NT1 (H1a; $p<0.0004$) and T2=RT1=MT2 (H1b; $p<0.0004$) within each participant, are both violated for all factors (see Appendix Table 3.17). Thus, workers in our sample report significant changes over time, when being asked to compare their current state with the past (both during T1 and T2).

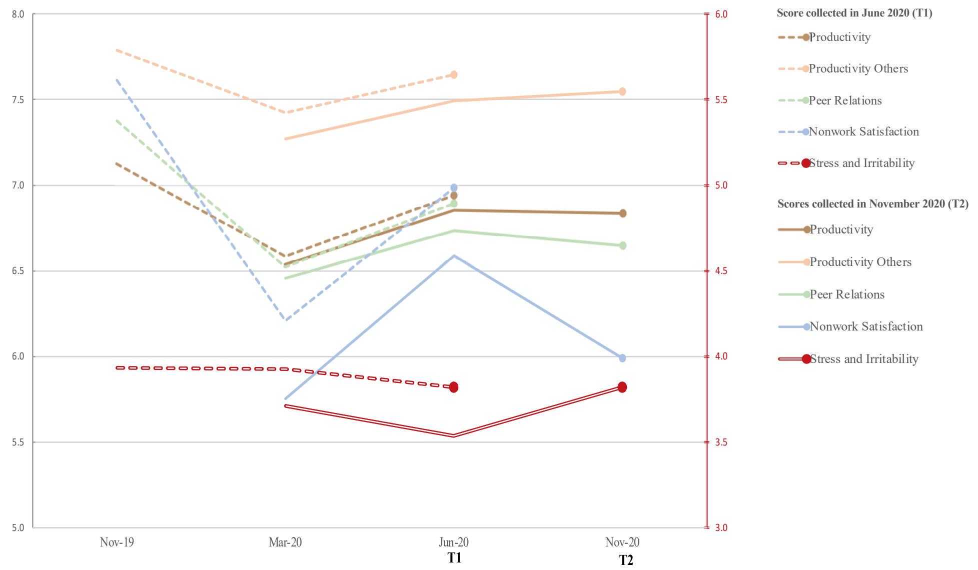
Table 3.8 Working from Home WPSQ scores for all time periods

	November 2019				March 2020						June 2020				November 2020			
	(1)		(2)		(3)		(4)		(5)		(6)		(7)		(8)		(9)	
	NT1		MT1		MT2		MT1=MT2		T1		RT1		T1=RT1		T2		T1=T2	
Variable	Mean	SD	Mean	SD	Mean	SD	*z*	*p*	Mean	SD	Mean	SD	*z*	*p*	Mean	SD	*z*	*p*
Productivity	7.12	0.96	6.58	1.24	6.54	1.41	1.45	.15	6.94	1.07	6.86	1.13	1.43	.15	6.84	1.28	0.86	0.37
Productivity by others	7.79	1.04	7.42	1.30	7.27	1.41	2.82	.00*	7.65	1.15	7.49	1.17	3.41	.00**	7.55	1.25	2.00	0.05
Stress and Irritability	3.93	1.55	3.93	1.52	3.71	1.56	4.90	.00***	3.82	1.46	3.54	1.47	6.80	.00***	3.82	1.63	1.38	0.17
Peer Relations	7.38	1.30	6.52	1.55	6.46	1.70	2.01	.04	6.89	1.41	6.73	1.48	3.91	.00***	6.65	1.59	4.88	.00***
Nonwork Satisfaction	7.61	1.09	6.21	1.48	5.75	1.76	6.76	.00***	6.98	1.17	6.59	1.37	6.58	.00***	5.99	1.59	15.84	.00***

Note. Z scores for pairwise Wilcoxon Signed Rank Test Significance is corrected by a Bonferroni multiple testing correction: *(.05).01, **(.01).002, and ***(.001).000

A selection of general trends within-factor and heterogeneity between-factors are noteworthy. For instance, productivity and peer relations display a comparable trend: starting high in November 2019, followed by a mild dip in March 2020, and recovering in between those scores in June and November 2020 (p<0.0004). Interestingly, stress and irritability remains relatively stable throughout the start of the pandemic (p>0.01), but dips only in June (p<0.0004). Non-work satisfaction shows the most movement over time (p<0.0004). Finally, how others think about their productive scores is consistently higher than all other factors, including their own productivity score. For a complete overview of all trends and post-hoc box plots, see Appendix Figure 3.5.

Figure 3.4 WPSQ Factor scores over time



Note. Figure 3.4 shows the trends of all WPSQ factors over the measurement periods. The point represents the current score, whereas each the dotted line represents the respective retrospective scores. For the factor Stress and Irritability, the right y-axis applies. For all other factors, the left y-axis applies. Standard deviation is not included, as these values merely indicate sample average trends. Significant difference will be estimated by non-parametric within-sample tests.

The first inspection of consistency between the two measurements (Hypothesis 2) shows that the participants' current sentiment in June (T1) seems largely unchanged in November (T2). Column 9 of Table 3.8 shows that only nonwork satisfaction and peer relations decreased significantly within subjects from T1 to T2 (Wilcoxon Signed Rank Test *z-score* of 4.880, p<0.0004; and 15.843, p<0.0004, respectively). This might indicate working from home fatigue: although productivity and stress are unchanged, both their personal life quality and peer relations are starting to suffer. For productivity, productivity by others, and stress, I find no significant difference which implies that hypothesis 2 cannot be rejected.

3.7.2 WPSQ factor Retrospection

Further careful inspection of Figure 3.4 suggests a trend in retrospective underestimation. Specifically, recollection scores from the T2 questionnaire (RT1 and MT2) are not only consistently lower than the scores from the first measurement, but also stable in this underestimation over time. First, all WPSQ factors recollections scores RT1 are lower than the actual score at T1 they are trying to recall (H3a). Accurate recollection should have resulted in similar scores of T1 and RT1. Column 7 of Table 3.8 tests this expectation. Within-subject non-parametric Wilcoxon signed rank test shows that all WPSQ factors, with the exception of productivity, indeed significantly underestimate their June recollection at November (Productivity by others: $z=3.41$, $p<.01$; for the remainder: $p<0.002$). Second, this trend appears to extend into the March recollection (H3b; see column 4 of Table 3.8 for the signed-rank test). This suggests that it's not mere recollection driving the underestimation in June: all June's recollections of March (MT1) scores are still (mostly significantly) higher can November's recollection of March (MT2). It is likely that the current (higher) satisfaction scores during June contaminate recollection in March similarly to how the lower satisfaction scores in November decrease the recollected satisfaction in June.

In order to investigate this recollective underestimation trend, I compare the objective scores at T1 with the recollection scores at RT1. Specifically, I compare the recollected score (RT1) with the actual score at that time (T1) as well as the current score during recollection (T2). The fact that both the recollected scores and the current scores at T2 are lower hints at a carry-over effect: the current score might influence the recollection more than the score it's supposed to recollect. As a first exploration, I compare the correlations of the WPSQ factors between the retrospective score, the target score, and the current score in the next section.

The retrospective scores (RT1) correlate with the target scores (T1), with correlation coefficient r between 0.46 and 0.63 (average of 0.54), shown in Appendix Table 3.15. This is surprisingly close to the correlations between the within-subject current scores (T1 and T2; 0.42 and 0.59, average of 0.53). The explanation of within-subject change over time does not apply to this retrospective assessment as long as participants are aware of their intrinsic change. Hence, the correlation between the retrospective score (RT1) and the target score (T1) is expected to be higher. Naturally, Appendix Table 3.15 also includes the correlations between retrospective score (RT1) and the current score (T2). Although these scores should correlate as the measure the same constructs at the same time, RT1 should correlate less with current T2 than with the actual target period of T1. However, the correlations are highest between RT1 and T2 (0.70 and

0.84, an average of 0.78; see Appendix Table 3.15). That only the RT1-T2 correlations can be considered strong (Schober et al., 2018), again signals that a recall bias might be present. In the next section I attempt to explore the retrospective scoring underlying mechanisms.

Table 3.9 shows a linear regression estimation for the retrospective productivity scores. The first model shows that the recollected productivity scores are significantly related to both the productivity at the time of recollection (T2), as well as the target productivity T1. Increased productivity at both moments increases retrospective productivity. Model 2 includes an interaction term of both variables. This model shows that, after controlling for the interaction, only the productivity during the recollection remains to significantly predict the recollection (α_2 = 0.25, p<0.05). If the recollection was accurate, not the current but the targeted productivity T1 should predict the recollection scores. This points towards a recall bias: recollection is more related to the current score than to the targeted score.

Table 3.9 WPSQ Productivity retrospective regressions

	(1) Retrospective Productivity RT1	(2) Retrospective Productivity RT1	(3) Δ Productivity T1- RT1	(4) |Δ|Productivity T1- RT1
Productivity T1	0.322***	0.0918		0.0986
	(0.0310)	(0.123)		(0.105)
Productivity T2	0.486***	0.253*	-0.176***	0.0506
	(0.0259)	(0.123)	(0.0277)	(0.105)
Productivity T1 * T2		0.0339		-0.0175
		(0.0176)		(0.0150)
Household Controls	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes
Job Controls	Yes	Yes	Yes	Yes
Constant	1.084***	2.654**	1.196**	45.50***
	(0.320)	(0.876)	(0.387)	(5.752)
Observations	772	772	772	772
R^2	0.603	0.605	0.065	0.383

Standard errors in parentheses. * p < 0.05, ** p < 0.01, *** p < 0.001

As a next step, models 3 and 4 explore the relationship between this carry-over effect and accuracy. It is unlikely that higher productivity would make participants more accurate in, yet model 3 shows that increased productivity decreases the difference between the targeted productivity and the recollected productivity (ΔProductivity: T1-RT1; β_1=-0,18, p<0.001). The systematic underestimation in recollection could make it possible that a carry-over effect of the more positive current state on recollection coincidentally also narrows the gap between recollection and targeted state. Therefore, model 4 predicts the retrospection bias using the absolute delta between the targeted and recollected state ($|\Delta|$ Productivity = |T1-RT1|). This absolute delta indicates the magnitude of the bias: if the absolute delta increases, the recall bias increases. This model definitively shows that, for productivity, the current state influences the recollection more than the targeted state. None of the productivity scores nor the interaction decreases the absolute delta significantly, implying that this effect does not increase the accuracy.

A similar pattern is discovered for all factors of the WPSQ.[43] Including an interaction term for T1 and T2 measures generally shows that either only the current score (T2) predicts the recollection, or that the coefficients of the T2 score are at least triple the magnitude that of the T1 target coefficients. This is unwanted and indicates a recall bias, since the target scores T1 should predict the recollection better than the current score. The raw delta between recollection and targeted score is subsequently consistently negatively related to the current score at T2, suggesting that a higher T2 score makes the recall bias smaller, or recollection more accurate. However, for all but one factor, the absolute delta estimation models show that higher T1 and/or T2 scores are either not associated with the recollection accuracy, or even negatively correlated with the recollection accuracy.

For one factor, non-work satisfaction, a deviating pattern is detected. Similar to all other factors,

Table 3.10 shows that higher current scores increase the recollection scores more than the targeted scores in model 2(α_2 = 0.50, p<0.001). Moreover, this recollection (raw delta, model 3) seems to shrink with increased satisfaction scores, implying increased accuracy (β_1=-0,24, p<0.001). However, contrasting all other factors, the absolute delta model 4 still indicates that increased satisfaction scores increase accuracy (γ_2=-20, p<0.001). Looking at Figure 3.4, one reason for this effect on non-work satisfaction could be the distinctly different pattern compared to the other factors. Nonwork satisfaction does not only have the largest delta, but is also the only factor that decreased in T2 compared to T1 (and RT1). As such, movement upwards is more likely to close the absolute recollection gap compared to factors whose T2 scores are closer to the RT1 (and T1). This interaction is explored by introducing a difference dummy between T1 and T2: if the previous pattern persists, the decreasing of the bias should only exist for participants who scored lower in T2 than in T1. If the nonwork satisfaction was higher in T2 than T1, and a carry-over effect exists, the increased accuracy found in model 4 is only the result

[43] The remaining WPSQ factors' retrospective regressions are shown in Appendix Table 3.18.

of an upward correction, and not of increased accuracy. Model 5 shows that, after introduction of this dummy, accuracy no longer increases for higher satisfaction scores. Although this model is admittedly simplistic, it confirms that the notion that increased non-work satisfaction improves retrospective accuracy is highly unlikely.

Table 3.10 WPSQ Nonwork Satisfaction retrospective regressions

	(1) Retrospective Nonwork Satisfaction RT1	(2) Retrospective Nonwork Satisfaction RT1	(3) Δ Nonwork Satisfaction T1- RT1	(4) |Δ| Nonwork Satisfaction T1- RT1	(4) |Δ| Nonwork Satisfaction T1- RT1
Nonwork Satisfaction T1	0.162***	0.121		0.0659	
	(0.0349)	(0.0901)		(0.0799)	
Nonwork Satisfaction T2	0.551***	0.499***	-0.238***	-0.203***	-0.169
	(0.0257)	(0.110)	(0.0295)	(0.0199)	(0.0902)
Nonwork Satisfaction T1*T2		0.00731		0.0111	
		(0.0151)		(0.0134)	
Dummy T1 versus T2					
T1>T2 (Decreased)					0.804
					(0.647)
T1<T2 (Increased)					0.0655
					(0.743)
T2 Decreased * Nonwork Satisfaction T2					-0.0952
					(0.0938)
T2 Increased * Nonwork Satisfaction T2					0.0597
					(0.105)
Household Controls	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes
Job Controls	Yes	Yes	Yes	Yes	Yes
Constant	1.555***	1.850*	2.384***	2.358***	2.059**
	(0.437)	(0.749)	(0.538)	(0.664)	(0.709)
Observations	772	772	772	772	772
R^2	0.528	0.528	0.111	0.170	0.185

Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Finally, I explore the discrepancy in retrospective scores for both T1 and T2 measures. The WPSQ scores for March are for both T1 as well as T2 based on retrospection. In order to explore the recall for these scores, I estimate the predictive value of the absolute delta of both current scores on T1 and T2 on the absolute delta of both retrospective scores of March (MT1 and MT2, respectively). Table 3.11 shows that, for each respective WPSQ factors, the difference between the current scores is predictive for the difference between the two retrospective scores (δ_1 ranging from 0.22 to 0.61, p<0.001). In other words, a larger change in a current state predicts a larger difference in the recollection of the same period. These findings, again, suggests the existence of a recall bias.

Table 3.11 WPSQ factors' recall regressions

	(1) |Δ| Productivity MarchT1 – MarchT2 (MT1 – MT2)	(2) |Δ| Productivity by others MarchT1 – MarchT2 (MT1 – MT2)	(3) |Δ| Peer Relations MarchT1– MarchT2 (MT1 – MT2)	(4) |Δ| Stress and Irritability MarchT1– MarchT2 (MT1 – MT2)	(5) |Δ| Nonwork Satisfaction MarchT1– MarchT2 (MT1 – MT2)
|Δ| Respective Factor T1– T2	0.220***	0.606***	0.531***	0.529***	0.392***
	(0.0344)	(0.0324)	(0.0333)	(0.0299)	(0.0340)
Household Controls	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes
Job Controls	Yes	Yes	Yes	Yes	Yes
Constant	0.503	0.872**	1.086**	0.108	1.116**
	(0.279)	(0.323)	(0.379)	(0.307)	(0.415)
Observations	772	727	741	772	772
R^2	0.079	0.340	0.279	0.312	0.171

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Overall, the results of this exploration into the recall bias suggest that there exists a recall bias in self-reported productivity and other work-related scores during the working-from-home shift. The current state predicts the recollection score consistently better than the targeted score. This confirms that productivity retrospection is as likely to be subject to recollection inaccuracy as health or other autobiographical domains (e.g. Bryant et al., 1989; Coughlin, 1990; Holmberg et al., 1996; Puente-Díaz & Cavazos-Arroyo, 2022; Weinstock et al., 1991). The fact that, in this sample, the current state scores are consistently lower than the targeted scores initially hinted towards a relation between the current state and accuracy. However, further inspection suggests that this effect is (at least partially) the result of a carry-over correction: on average, higher current state scores lead to higher recollection scores, decreasing the underestimation in recollection. In absolute terms, the recall often increases with higher current state scores. As such, these results show evidence for a carry-over effect of the current state into the recollection scores, resulting in a recall bias. The retrospective assessment of productivity in a largely dynamic environment is strongly influenced by the current state. Therefore, the absolute objective value of retrospective reports of productivity is relatively low.

3.8 General Conclusion

Accurately monitoring work-related productivity and stress has become increasingly important as a result of the pandemic-driven surge in work from home (Aksoy et al., 2022; Barrero et al., 2021; Bloom et al., 2022; Etheridge et al., 2020). This paper sets out to reconstruct and validate an easy-to-administer, self-reported questionnaire that incorporates all relevant facets of work productivity and stress. A factor analysis results in a restructured questionnaire, the Work Productivity and Stress Questionnaire (WPSQ), which identifies five factors: productivity, productivity by others, peer relations, nonwork satisfaction, and stress and irritability. I show that the redeveloped WPSQ factors outperform previously constructed factors on a large sample.

The WPSQ supports employers in accurately assessing the current state of their employees using a validated and reliable tool. This tool can be used to assess the general productivity trend, but can also be used to evaluate a change in policy or, as I show in this paper, a change in the work environment (e.g. Chapter 4). The multidimensional factors of this questionnaire furthermore extend the assessment of productivity and include how people perceive their peer relations, their nonwork satisfaction, and think others perceive their productivity. Additionally, the stress and irritability factor does not only inform about stress itself, but is also closely related to burnout propensity, and absenteeism. High scores on this factor clearly indicate that further inspection into these two highly potent domains is warranted. Finally, I show that single-item scales could function as alternatives for some WPSQ factors, but only when done with caution and only when brevity demands it.

When the WPSQ is applied to working from home I show exploratory trends over time in a volatile, pandemic-ridden context. A recollection (or recall) bias is often overlooked in the productivity domain, yet inherent to self-reported (autobiographical) recollection of questionnaire scores over time (Coughlin, 1990; Drews & Greeland, 1990; Puente-Díaz & Cavazos-Arroyo, 2022; Schacter, 1999). In this paper, I confirm that the retrospective scores are heavily influenced by a recall bias: the current state colors the recollection of the past. With a regression approach, I further debunk the initial observation suggesting that increased current scores increase accuracy. The observation that the recollection bias is more related to the current state than the state we try to remember is not trivial. Based on these results, I conclude that the retrospective scores for the start of the pandemic are, although probably related to the real scores, influenced by the state during measurement. The retrospective evaluation of working from home during the pandemic, for instance, could be biased as a result. This paper suggests that the interpretation of retrospective work-related scores should be done with caution. Retrospective scoring is subject to a recall bias and has the potential to reflect a non-existing trend (Baars & Franklin, 2003; Wilson & Ross, 2003).

The following limitations should be acknowledged. First, I confirm the strength of the WPSQ factors allocation with multiple measures in a large sample. However, it is possible that this sample is not sufficiently heterogeneous to validate the factor allocations to a population (Boateng et al., 2018). It is recommended that the performance of these factors is investigated when this questionnaire is deployed, to ensure a strong fit of factors to a different target audience. Furthermore, I utilize single-scale items as proxies for absenteeism and a shortened burnout scale. More extended measures or different proxies should confirm the conclusion I draw from these specific measures.

Second, the recall bias is investigated by comparing the recollection score with the targeted self-reported score. Although the aim of this investigation is to show the accuracy of recollection, it must be noted that I technically investigate the difference between two measures of self-report. Ideally, I would also compare the self-reported scores with realized productivity. Although the relevance of this questionnaire is emphasized by the fact that there is limited information available on the realized productivity, future research should focus on investigating the accuracy gap between realized and self-reported scores in the work domain.

Finally, the results suggest that recollection accuracy is influenced by recall bias, yet I cannot infer which mental motivation drives this bias. I discuss multiple factors that have been found to drive a recall bias, but the collected data does not enable me to conclude which mechanism underpins these inaccuracies. For instance, the recall bias could be a combination of inaccurate memory and anchoring, where blurred past is informed by the current state. This could be both mental as well as technical anchoring: a person could use their current sentiment as anchor, but also could use their stated answer on paper as anchor. In the latter, a recall bias is likely to be less prevalent in an alternative survey style, such as qualitative interviews. The observation that recall bias exists in this domain is valuable, yet it could also be that this observation is for instance driven by a situational mechanism related to the pandemic. Although these results are in line with an array of research in different domains, the generalization of these results would improve by more future research focus on the driver of this bias in this specific domain.

3.9 Appendix

Table 3.12 PCA Eigenvalues of the Work Satisfaction and Productivity Questionnaire (WPSQ)

	Eigenvalue	Difference	Proportion	Cumulative
Comp1	9.53	6.31	0.32	0.32
Comp2	3.22	1.05	0.11	0.42
Comp3	2.16	0.64	0.07	0.50
Comp4	1.52	0.30	0.05	0.55
Comp5	1.22	0.24	0.04	0.59
Comp6	0.98	0.01	0.03	0.62
Comp7	0.98	0.07	0.03	0.65
Comp8	0.91	0.09	0.03	0.68
Comp9	0.82	0.10	0.03	0.71
Comp10	0.73	0.03	0.02	0.74
Comp11	0.70	0.05	0.02	0.76
Comp12	0.65	0.03	0.02	0.78
Comp13	0.62	0.03	0.02	0.80
Comp14	0.59	0.02	0.02	0.82
Comp15	0.57	0.02	0.02	0.84
Comp16	0.55	0.03	0.02	0.86
Comp17	0.52	0.08	0.02	0.88
Comp18	0.44	0.01	0.01	0.89
Comp19	0.43	0.05	0.01	0.91
Comp20	0.39	0.03	0.01	0.92
Comp21	0.36	0.05	0.01	0.93
Comp22	0.31	0.02	0.01	0.94
Comp23	0.29	0.02	0.01	0.95
Comp24	0.28	0.04	0.01	0.96
Comp25	0.24	0.01	0.01	0.97
Comp26	0.23	0.01	0.01	0.98
Comp27	0.22	0.02	0.01	0.98
Comp28	0.20	0.03	0.01	0.99
Comp29	0.17	0.02	0.01	0.99
Comp30	0.15	.	0.01	1.00

Table 3.13 Factor Comparison and Item Allocations between the Work Productivity and Stress Questionnaire (WPSQ) and the Health Work Questionnaire (HWQ)

	Work Satisfaction and Productivity Questionnaire (WPSQ)			Health Work Questionnaire (HWQ)		
	ID	Factor	Factor Label	ID	Factor	Factor Label
Your view of your efficiency this week	01	1	Productivity	12a	1	Productivity
Your view of your quality this week	04	1	Productivity	13a	1	Productivity
Your amount of work this week	07	1	Productivity	14a	1	Productivity
Rate highest level of efficiency	10	1	Productivity	15	1	Productivity
Rate lowest level of efficiency	11	1	Productivity	16	1	Productivity
Frequency of boredom at work	13	1	dropped	21	2	Concentration/Focus
How satisfied with your job?	19	1	Productivity	5	4	Work Satisfaction
How satisfied with work environment?	21	1	Productivity	2	5	Work Satisfaction
How rewarding was work in past week?	22	1	Productivity	3	5	Work Satisfaction
Frequency of low concentrating at work	14	1	dropped	22	2	Concentration/Focus
Coworkers' view of your efficiency	03	2	Productivity by others	12c	1	Productivity
Supervisor's view of your quality	05	2	Productivity by others	13b	1	Productivity
Coworkers' view of your quality	06	2	Productivity by others	13c	1	Productivity
Supervisor's view of your amount of work	08	2	Productivity by others	14b	1	Productivity
Coworkers' view of your amount of work	09	2	Productivity by others	14c	1	Productivity
Supervisor's view of your efficiency	02	2	Productivity by others	12b	1	Productivity
Frequency of restlessness at work	12	3	Stress and Irritability	20	2	Concentration/Focus
Frequency of annoyance with coworkers	25	3	Stress and Irritability	17	6	Impatience/irritability
Frequency of impatience with coworkers	26	3	Stress and Irritability	18	6	Impatience/irritability
Frequency of conflicts with coworkers	27	3	Stress and Irritability	19	6	Impatience/irritability
How stressed are you?	28	3	Stress and Irritability	1	-	dropped
Frequency of unable to complete tasks [2]	30	3	Stress and Irritability	23	-	dropped
Frequency of being too exhausted to do work	15	3	Stress and Irritability	24	2	Concentration/Focus
How satisfied with supervisor relationship?	16	4	Peer Relations	8	3	Supervisor Relations
How easy to communicate with supervisor?	17	4	Peer Relations	10	3	Supervisor Relations
How satisfied with coworker relationship?	24	4	Peer Relations	7	5	Work Satisfaction
Experienced control on work	29	-	dropped	9	-	dropped
How rewarding was personal life?	18	5	Nonwork Satisfaction	4	4	Nonwork Satisfaction
How easy to communicate with family/friends?	20	5	Nonwork Satisfaction	11	4	Nonwork Satisfaction
How satisfied with relationships?	23	5	Nonwork Satisfaction	6	5	Nonwork Satisfaction

Table 3.14 Inter-item correlations of all WPSQ factors

Variables	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10
(1) WPSQ 1	1									
(2) WPSQ 4	0.65	1								
(3) WPSQ 7	0.66	0.58	1							
(4) WPSQ 10	0.62	0.56	0.58	1						
(5) WPSQ 11	0.40	0.33	0.38	0.34	1					
(6) WPSQ 13	0.42	0.37	0.38	0.39	0.29	1				
(7) WPSQ 19	0.52	0.48	0.42	0.48	0.36	0.51	1			
(8) WPSQ 21	0.31	0.30	0.25	0.29	0.20	0.24	0.38	1		
(9) WPSQ 22	0.52	0.45	0.43	0.49	0.34	0.53	0.79	0.42	1	
(10) WPSQ 14	0.45	0.39	0.35	0.39	0.35	0.63	0.47	0.29	0.47	1

3.14.1 Inter-item correlation matrix 'Productivity' (factor 1)

3.14.2 Inter-item correlation matrix 'Productivity by others' (factor 2)

Variables	-1	-2	-3	-4	-5	-6
(1) WPSQ 3	1					
(2) WPSQ 5	0.53	1				
(3) WPSQ 6	0.66	0.73	1			
(4) WPSQ 8	0.49	0.65	0.54	1		
(5) WPSQ 9	0.63	0.53	0.66	0.70	1	
(6) WPSQ 2	0.65	0.67	0.50	0.61	0.48	1

3.14.3 Inter-item correlation matrix 'Stress and Irritability' (factor 3)

Variables	-1	-2	-3	-4	-5	-6	-7
(1) WPSQ 12	1						
(2) WPSQ 25	0.42	1					
(3) WPSQ 26	0.42	0.71	1				
(4) WPSQ 27	0.32	0.55	0.57	1			
(5) WPSQ 28	0.51	0.37	0.36	0.27	1		
(6) WPSQ 30	0.32	0.29	0.29	0.31	0.32	1	
(7) WPSQ 15	0.48	0.38	0.35	0.34	0.45	0.48	1

3.14.4 Inter-item correlation matrix 'Peer Relations' (factor 4)

Variables	-1	-2	-3
(1) WPSQ 16	1		
(2) WPSQ 17	0.68	1	
(3) WPSQ 24	0.54	0.33	1

3.14.5 Inter-item correlation matrix 'Nonwork Satisfaction' (factor 5)

Variables	-1	-2	-3
(1) WPSQ 18	1		
(2) WPSQ 20	0.39	1	
(3) WPSQ 23	0.48	0.63	1

Table 3.15 Correlations between WPSQ factors scores T1 – T2 – Retrospect

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
(1) Productivity T2	1														
(2) Productivity T1	0.59	1													
(3) Productivity RT1	0.72	0.63	1												
(4) Productivity by others T2	0.61	0.41	0.50	1											
(5) Productivity by others T1	0.34	0.56	0.37	0.42	1										
(6) Productivity by others RT1	0.45	0.41	0.59	0.84	0.45	1									
(7) Stress and irritability T2	-0.54	-0.40	-0.40	-0.26	-0.19	-0.16	1								
(8) Stress and irritability T1	-0.32	-0.48	-0.31	-0.19	-0.23	-0.17	0.59	1							
(9) Stress and irritability RT1	-0.38	-0.39	-0.46	-0.22	-0.23	-0.22	0.83	0.59	1						
(10) Peer Relations T2	0.49	0.39	0.40	0.39	0.32	0.31	-0.34	-0.21	-0.28	1					
(11) Peer Relations T1	0.30	0.47	0.31	0.28	0.41	0.28	-0.23	-0.28	-0.25	0.57	1				
(12) Peer Relations RT1	0.33	0.38	0.49	0.31	0.34	0.39	-0.23	-0.20	-0.28	0.79	0.59	1			
(13) Nonwork Relations T2	0.36	0.27	0.29	0.24	0.08	0.19	-0.25	-0.12	-0.19	0.48	0.28	0.39	1		
(14) Nonwork Relations T1	0.30	0.40	0.30	0.25	0.19	0.22	-0.25	-0.26	-0.25	0.37	0.49	0.38	0.50	1	
(15) Nonwork Relations RT1	0.23	0.23	0.37	0.19	0.10	0.21	-0.16	-0.10	-0.24	0.37	0.25	0.43	0.70	0.46	1

Note. Table 3.15 shows the correlations between all WPSQ factor November (T2), June (T1), and retrospective June (RT1) scores. Ideally, T1 would correlate with T2 with stable sentiment, and regardless of sentiment change, RT1 would correlate higher with T1 (target score) than with T2 (the current score). Therefore, the Table coloring indicates the consistency correlation (T1 and T2) in green, the targeted retrospective correlation grey (RT1 and T1), and the carryover correlation (RT1 and T2) in red.

Table 3.16 Convergent and Discriminant Validity Productivity and Stress

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) Productivity (WPSQ)	1							
(2) How Productive at work	0.73	1						
(3) How Satisfied with work	0.70	0.56	1					
(4) Stress (WPSQ)	-0.55	-0.37	-0.48	1				
(5) Burnout Propensity (MBI)	-0.50	-0.31	-0.46	0.71	1			
(6) Breaktime	-0.11	-0.13	-0.03	-0.08	-0.10	1		
(7) Days off work	0.00	-0.06	0.03	-0.11	-0.14	0.30	1	
(8) Sickdays	-0.21	-0.17	-0.13	0.21	0.28	0.07	0.13	1

Table 3.17 Extended Mantel-Haenszel Stratified Test of Association – Friedman's Test

	June 2020 measurement								November 2020 measurement							
	(1)		(2)		(3)		(4)		(5)		(6)		(7)		(8)	
	NT1		MT1		T1				MT2		RT1		T2			
Variable	Mean	SD	Mean	SD	Mean	SD	Q	p	Mean	SD	Mean	SD	Mean	SD	Q	p
Productivity	7.12	0.96	6.58	1.24	6.94	1.07	202.20	.00***	6.54	1.41	6.86	1.13	6.84	1.28	62.64	.00***
Productivity by others	7.79	1.04	7.42	1.30	7.65	1.15	159.22	.00***	7.27	1.41	7.49	1.17	7.55	1.25	50.71	.00***
Stress and Irritability	3.93	1.55	3.93	1.52	3.82	1.46	26.76	.00***	3.71	1.56	3.54	1.47	3.82	1.63	86.74	.00***
Peer Relations	7.38	1.30	6.52	1.55	6.89	1.41	591.71	.00***	6.46	1.70	6.73	1.48	6.65	1.59	66.97	.00***
Nonwork Satisfaction	7.61	1.09	6.21	1.48	6.98	1.17	773.05	.00***	5.75	1.76	6.59	1.37	5.99	1.59	370.67	.00***

Note. Q scores for non-parametric repeated-measure pairwise ANOVA alternative, Friedman Test. Significance is corrected by a Bonferroni multiple testing correction: *(.05).01, **(.01).002, and ***(.001).0004

Table 3.18 WPSQ Factors' retrospective regressions

3.18.1 WPSQ Productivity's retrospective regression

	(1) Retrospective Productivity RT1	(2) Retrospective Productivity RT1	(3) Δ Productivity T1-RT1	(4) |Δ| Productivity T1-RT1
Productivity T1	0.322*** (0.0310)	0.0918 (0.123)		0.0986 (0.105)
Productivity T2	0.486*** (0.0259)	0.253* (0.123)	-0.176*** (0.0277)	0.0506 (0.105)
Productivity T1 * T2		0.0339 (0.0176)		-0.0175 (0.0150)
Household Controls	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes
Job Controls	Yes	Yes	Yes	Yes
Constant	1.084*** (0.320)	2.654** (0.876)	1.196** (0.387)	45.50*** (5.752)
Observations	772	772	772	772
R^2	0.603	0.605	0.065	0.383

Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

3.18.2 WPSQ Stress (and Irritability)'s retrospective regression

	(1) Retrospective Stress RT1	(2) Retrospective Stress RT1	(3) Δ Stress T1- RT1	(4) |Δ| Stress T1- RT1
Stress T1	0.157*** (0.0256)	0.267*** (0.0514)		0.611*** (0.0519)
Stress T2	0.670*** (0.0226)	0.777*** (0.0492)	-0.255*** (0.0292)	0.386*** (0.0497)
Stress T1 * T2		-0.0275* (0.0112)		-0.106*** (0.0113)
Household Controls	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes
Job Controls	Yes	Yes	Yes	Yes
Constant	0.0262 (0.340)	-0.363 (0.374)	1.284* (0.527)	-1.506*** (0.378)
Observations	772	772	772	772
R^2	0.707	0.710	0.114	0.199

Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

3.18.3 WPSQ Peer Relations retrospective regression

	(1) Retrospective Peer Relations RT1	(2) Retrospective Peer Relations RT1	(3) Δ Peer Relations T1-RT1	(4) |Δ| Peer Relations T1-RT1
Peer Relations T1	0.209***	0.164*		0.197**
	(0.0275)	(0.0712)		(0.0736)
Peer Relations T2	0.651***	0.601***	-0.247***	0.225**
	(0.0251)	(0.0770)	(0.0304)	(0.0796)
Peer Relations T1 * T2		0.00752		-0.0431***
		(0.0108)		(0.0112)
Household Controls	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes
Job Controls	Yes	Yes	Yes	Yes
Constant	1.099**	1.398*	2.628***	0.717
	(0.398)	(0.587)	(0.550)	(0.606)
Observations	741	741	741	741
R^2	0.666	0.666	0.115	0.072

Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

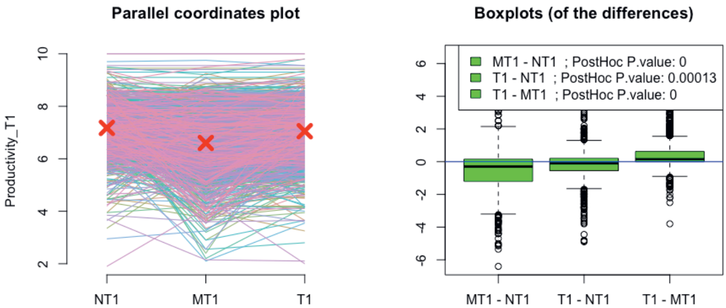
3.18.4 WPSQ Productivity by others retrospective regression

	(1) Retrospective Productivity by others RT1	(2) Retrospective Productivity by others RT1	(3) Δ Productivity by others T1- RT1	(4) |Δ| Productivity by others T1- RT1
Productivity by others T1	0.124***	0.162		1.239***
	(0.0220)	(0.103)		(0.129)
Productivity by others T2	0.730***	0.768***	-0.406***	1.220***
	(0.0204)	(0.103)	(0.0336)	(0.129)
Productivity by others T1 * T2		-0.00519		-0.182***
		(0.0136)		(0.0170)
Household Controls	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes
Job Controls	Yes	Yes	Yes	Yes
Constant	1.370***	1.088	2.675***	-7.347***
	(0.285)	(0.792)	(0.479)	(0.993)
Observations	727	727	727	727
R^2	0.733	0.733	0.209	0.216

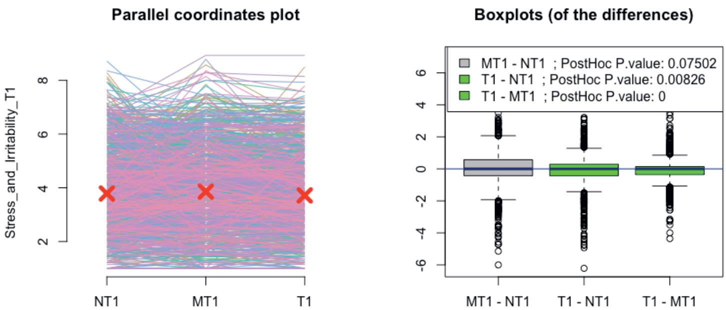
Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure 3.5 Friedman's Test Post Hoc Coordinates plots and differences box plot

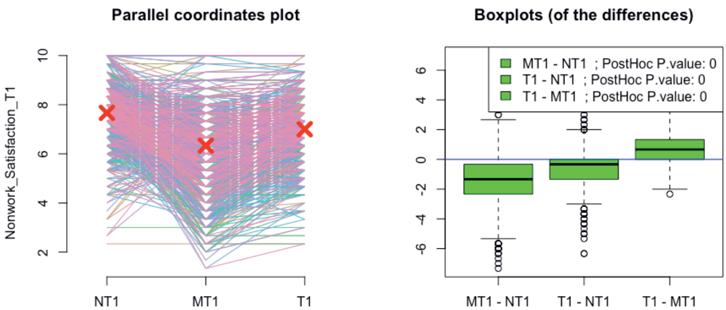
3.5.1 Coordinates plots and differences box plot Productivity for measurement T1



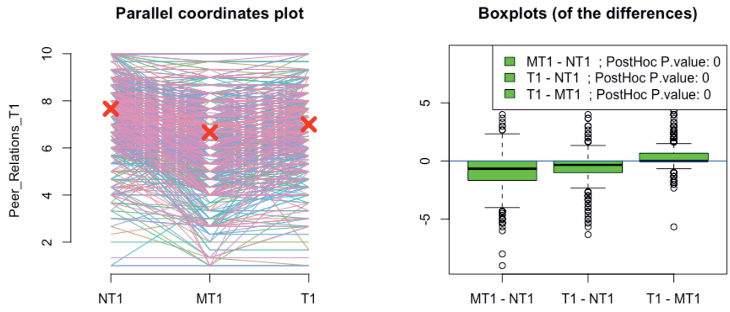
3.5.2 Coordinates plots and differences box plot Stress and Irritability for measurement T1



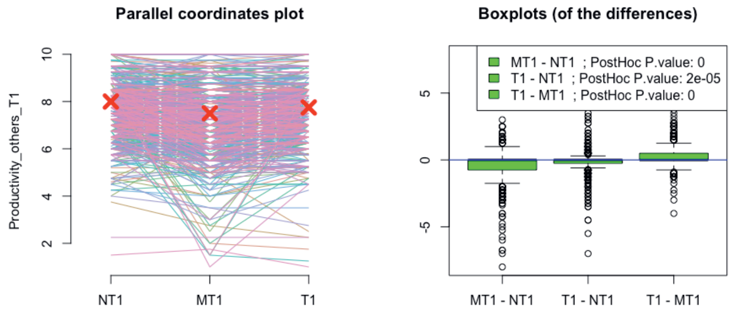
3.5.3 Coordinates plots and differences box plot Nonwork Satisfaction for measurement T1



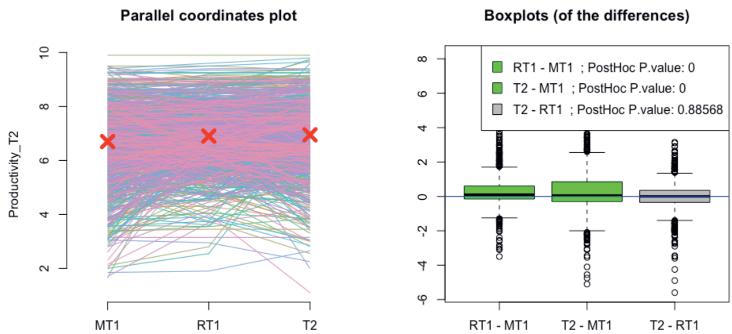
3.5.4 Coordinates plots and differences box plot Peer Relations for measurement T1



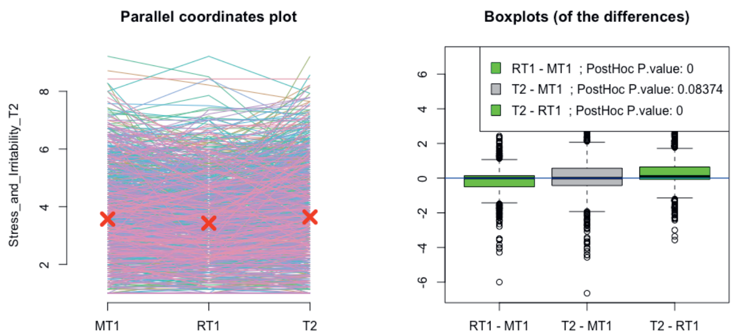
3.5.5 Coordinates plots and differences box plot Productivity by others for measurement T1



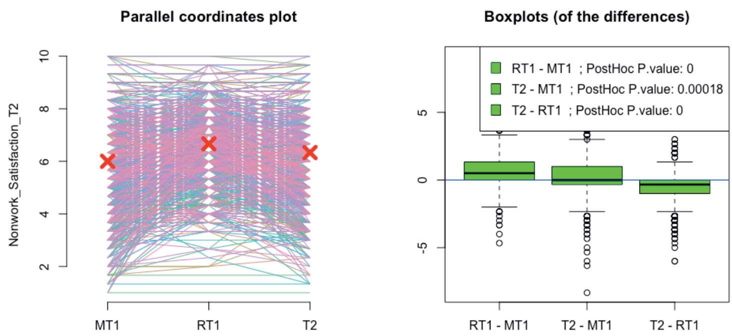
3.5.6 Coordinates plots and differences box plot Productivity for measurement T2



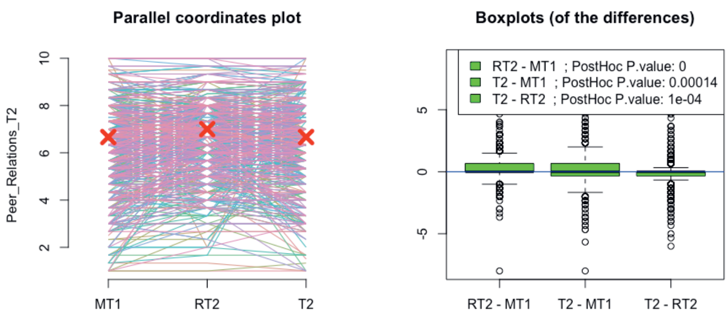
3.5.7 Coordinates plots and differences box plot Stress and Irritability for measurement T2



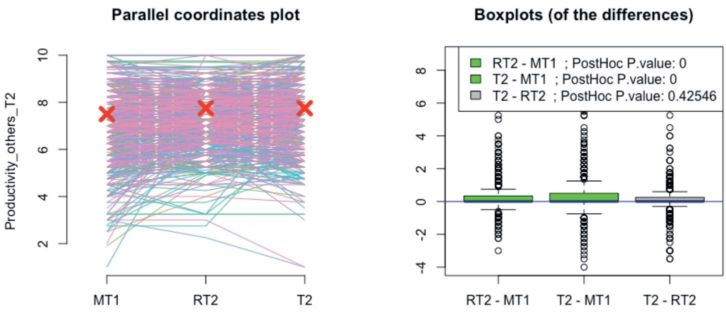
3.5.8 Coordinates plots and differences box plot Nonwork Satisfaction for measurement T2



3.5.9 Coordinates plots and differences box plot Peer Relations for measurement T2



3.5.10 Coordinates plots and differences box plot Productivity by others for measurement T2



Chapter 4

Does working from home work? That depends on the home.*

4.1 Introduction

The COVID-19 pandemic, in combination with recent technological advancements, has quickly elevated the status of working from home (WFH) from "occasionally" to "the new normal"(Barrero et al., 2021). Earlier uncertainty about the quantity and quantity of work produced at home had hampered large-scale corporate acceptance (Swisher, 2013, Belanger, 1999). However, these doubts were simply overruled by the COVID-19 pandemic, which forced the majority of knowledge-based employees to work online. Negative stigmas previously related to working from home diminished drastically (Barrero et al., 2021), at least temporarily. In addition, prior technological complications were quickly overcome following a pandemic-driven surge in technological innovations, such as the advent of Teams and Zoom calls. This involuntary litmus test pushed working from home out of its infancy. However, what has gained limited attention in explaining the success of working from home is the physical climate in which daily work takes place – the home office.

This paper studies individual differences in working from home productivity, as it relates to the heterogeneity of the physical home office environment and the satisfaction of employees. Using a series of survey administered at various points throughout the pandemic, we find that the self-reported productivity of a large sample of Dutch home workers was lower at home compared to previously at the work office, before the COVID-19 pandemic. Generally, both office hardware satisfaction (e.g. screen, Wi-Fi, and desk; on average preferred at the work office) and office indoor environment satisfaction (e.g. air quality, temperature, and noise; on average preferred at the home office) directly affect working from home productivity, burnout propensity, and willingness to continue to work from home.

We subsequently study individual behavior that could influence the indoor environment at home. We show that the degree of ventilation (the percentage of time the home office is ventilated) during working hours affects self-reported productivity, burnout propensity, and willingness to continue to work from home. This effect is fully mediated by the degree of self-reported satisfaction with the physical environment. We show that increasing ventilation is associated with a general increase in satisfaction scores. This is true even for satisfaction with hardware, which is, intuitively, unrelated to ventilation. Furthermore, the self-reported satisfaction that is most strongly related to ventilation (e.g. air quality satisfaction) was the only factor initially lacking a strong relationship with productivity. Thus, solely based on self-report analysis, ventilation would have been an unlikely factor considered to improve the success of work from home.

The main contribution of this paper is to show that the physical climate influences employee productivity and satisfaction when working from home. Specifically, we not only connect the outcomes of work from home to self-reported climate satisfaction, but also to indoor environment-influencing behavior. The move from the work office to the home office needs to be combined with careful design and investment in the quality of the office *and* its climate. Failure to do so is not only likely to be associated with decreased productivity, but also decreased willingness to work from home, and increased burnout propensity. The physical climate is a determining factor in successful work from home prolongation. Additionally, our results also suggest it is crucial that the physical environment is objectively measured, as merely collecting self-reported satisfaction scores might paint an incomplete or even incorrect picture. As such, this paper reaffirms that the effect of a healthy indoor climate affects productivity, related to previous research that shows significant health effects of indoor climate (J. G. Allen et al., 2016; Bluyssen, 2012; MacNaughton et al., 2017; Palacios et al., 2020).

The remainder of our paper is organized as follows. In the next section, we provide an overview of the current state of the literature and the background of working from home. Section 3 provides an overview of the metrics we included in our survey design. In section 4 we describe our sample and also explore the difference in reported scores for the home office compared to the work office. Section 5 discusses the model used to estimate our findings. The results of our regression models, as well as our mediation models, are presented in section 6. Finally, section 7 concludes.

4.2 Background literature

The rising popularity of work from home in the last decade has been well-reported: a recent report by buffer.com (Griffis, 2021) amongst 2,300 workers showed that over 97% would like to continue to work from home, at least partially. Employees are, on average, willing to take a 5% pay cut for 2-3 days of work from home (Aksoy et al., 2022). Employees working from home report being as productive as they were at the office before the pandemic (Etheridge et al., 2020). These positive experiences have led to the prediction that, after the pandemic, 20% of all office work will be carried out from home. This continuation of work from home is expected to boost productivity by almost 5%, although largely unobservable by standard measures, as it stems mainly from a reduction in commuting (Barrero et al., 2021).

Working from home has clear advantages as well as disadvantages for both job performance and human health and well-being. Multiple studies show positive effects on job satisfaction and turnover intent (T. D. Allen et al., 2015; Bloom et al., 2015; Gajendran & Harrison, 2007). Bloom et al. (2014) report that work from home leads to less commuting and fewer distractions. In addition, exhaustion leading to burnout is negatively related to work from home (Sardeshmukh et al., 2012). Perceived autonomy seems to be one of the main drivers of these positive effects: the degree to which employees can choose a location and time to work, independently of their supervisors, both predict the intensity of working from home, as well as job performance, mental burnout, and job dedication, even during the pandemic (Bloom et al., 2014; Gajendran et al., 2015; Sardeshmukh et al., 2012; B. Wang et al., 2021).

More recently, Bloom et al. (2022) found only modest self-reported and realized productivity increases for work from home during COVID-19, whereas others identified productivity decreases for those who did not work from home before the pandemic, suggesting selection bias in previous studies (Morikawa, 2022). Moreover, output assessments amongst ICT workers suggest productivity actually drops at home (Gibbs et al., 2021). In the past, the positive relationship between work from home intensity and productivity has repeatedly been found to be non-linear. According to Golden & Vega (2005), the optimal intensity to work from home is limited to about 16 hours per week. Work from home more and job satisfaction and performance could decrease. A survey by State of the Work in 2022 found that, amongst 2,000 respondents, 45% think career growth will be at risk with increased work from home (Griffis, 2022). Unsurprisingly, it is coworkers' relationships that suffer most from working from home, leading to professional isolation, which in turn has the potential to escalate into decreased performance and turnover intent (Gajendran & Harrison, 2007). Offline or online communication could mitigate these negative effects, but possibly only partially (Golden et al., 2008; B. Wang et al., 2021). For instance, Yang et al. (2022) find that firm-wide remote work inevitably lowers communication quality, as less communication leads to a worsening of information sharing.

Beyond having implications for coworker relationships, work from home may bring new interpersonal problems to light. Felstead & Henseke (2017) suggest that homeworkers are burdened by the "social exchange theory": they work harder, longer, and work unpaid hours in order to justify their freedom to work from a preferred location. Workers thus (over)compensate for the perception that they might work less when not being observed. The resulting work exhaustion has the ability to offset all positive effects of work from home on productivity, and may even lead to burnout symptoms (Golden, 2006). In addition, research shows that people working from home find it hard to detach from work, disrupting their established work-life balance (B. Wang et al., 2021). Interestingly, the work-family conflict was previously considered to decrease with work from home, supposedly due to increased autonomy (Gajendran & Harrison, 2007). The current perception of work from home having a negative impact on work-life balance could also be a pandemic-specific challenge.

Although overall perspectives on work from home vary, it is also important that beyond the average effects, substantial heterogeneity has been documented across jobs and individuals. To our knowledge, this heterogeneity has solely been explained by work and personal characteristics. For instance, the degree to which a job is suitable for work-from-home strongly predicts productivity (Etheridge et al., 2020). A job previously executed from behind a desk (e.g. financial services) is more easily shifted to a home office as compared to a manual, labor-orientated occupation. A heavy workload and the degree of monitoring by supervisors also negatively impact the work effectiveness from home (B. Wang et al., 2021). Jobs that have high levels of interdependence with colleagues, or are outcome-oriented, suffer when work from home intensity increases (Virick et al., 2010). Overall, limited support and inadequate feedback by the employer mitigate the positive effects of work from home (Sardeshmukh et al., 2012; B. Wang et al., 2021).

At the individual level, self-discipline seems to be a key factor for an efficient work from home (B. Wang et al., 2021). The degree to which an individual is able to ignore distractions that they do not have at the office is important, especially without the same level of social control by co-workers. Additionally, women seem to suffer more from work from home as compared to men (Etheridge et al., 2020). Women state their job to be less suitable for work from home in general and the presence of children affects work from home productivity for women more negatively as compared to men (Adams-Prassl et al., 2020, 2022; Manzo & Minello, 2020). Finally, the pandemic showed that young workers seem to appreciate work from home more, and opted for work from home more often as compared to older workers (Brynjolfsson et al., 2020). These results, however, are not stable per se. Another study shows opposite results, where both women and older workers reported being more productive when working from home (Awada et al., 2021).

What has gained limited attention in explaining individual differences in work from home satisfaction is the physical climate in which the daily work takes place. The COVID-19 pandemic has led to increased attention to the effect of indoor space on pathogen spreading. Specifically, ventilation has become the spearhead combating the airborne spreading of the COVID-19 virus at public and private indoor gatherings (J. G. Allen & Ibrahim, 2021; Somsen et al., 2020). The attention to air quality reinforces an existing trend where workplace quality is more and more important. In the office, employers aim to facilitate a healthy and comfortable work environment for employees, with the goal of promoting productivity (Al Horr et al., 2016; J. G. Allen et al., 2016; MacNaughton et al., 2017). Bad air and light quality, temperature, and noise have all been shown to negatively affect productivity and increase sick building symptoms, such as headaches in the office (Cedeño Laurent et al., 2018; L. Fang et al., 2004; Palacios et al., 2020; Wong, 2004). Hence, ergonomics, temperature, and noise pollution are all considered by modern employers in order to minimize interference with comfort and wellbeing in the office (Coovert & Thompson, 2003).

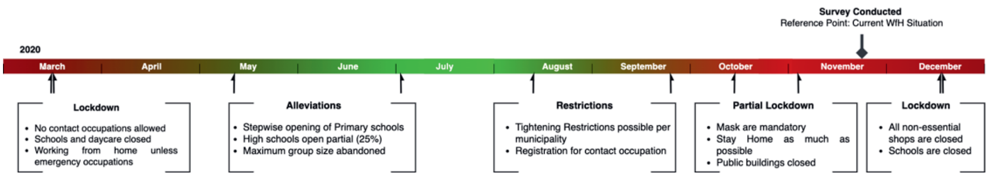
For the move to the home office, a trade-off is to be expected. On the one hand, suboptimal ergonomics at home are not as easily mitigated (Davis et al., 2020). For instance, not having a dedicated office negatively influences productivity at home (Awada et al., 2021). On the other hand, research suggests that controlling the thermostat at home might benefit the work from home satisfaction (Chang & Kajackaite, 2019; Stroom, Kok, et al., 2021). Looking at indoor environmental quality more broadly, Tahmasebi et al. (2021) show that people working at home during the pandemic close their windows more often as compared to before the lockdown. Combined with CO_2 data, they conclude that working from home often leads to worse indoor air quality. Generally, the professionalism or quality of the environment might suffer, but the workers' experienced control over these conditions at home might increase.

4.3 Work from home survey

We surveyed 1,002 Dutch individuals via the Flycatcher panel. Flycatcher is an academically-orientated research organization that established a high-quality panel representing the Dutch population[44]. Flycatcher randomly selected participants from their panel for an online survey, where participation was reimbursed. For the purpose of our research, we included only office workers (with a minimum age of 18 years old), who worked at least part-time from home at the time of the survey. People without work, previously without work, or working exclusively from the office were excluded from our sample.[45]

The first data collection took place in November 2020. At that time, the Netherlands had been in some form of lockdown for over 8 months due to the COVID-19 pandemic, since March 2020. Figure 4.1 provides a colorized overview of the timing data collection with respect to the Dutch restriction's development. Working from home was strongly recommended by the government, with some exceptions including workers in healthcare and essential shops. During this time, employers were not allowed to force their employees to come to the office. All other social activities were severely limited. A selection of questions was answered based on two moments in time: "current" (i.e. November 2020, working from home) and one year ago (working from the office).

Figure 4.1 Working from Home and Survey Timeline



It is relevant to point out that we utilize the COVID-19 restrictions to eliminate the selection problems hampering previous research. Before the pandemic-related restrictions, the success and satisfaction of working from home could potentially be explained by self-selection

[44] See for example (Bults et al., 2011; Poelman et al., 2021; Smeets et al., 2015)) for well published studies using the Flycatcher panel.
[45] The research setup was reviewed and approved by Maastricht University's Ethical Review Committee Inner City Faculties (ERCIC_195_09_06_2020).

following the request to (voluntarily) move to work from home. Inherent intrinsic motivation, personal characteristics, and ability to adjust to the physical environment could all be omitted factors in that request. From a company perspective, those previously offering WFH in the first place were likely by design willing to enable the shift to the home office, and must have had job characteristics which fit (at least partially) WFH. Due to the pandemic, this selection bias susceptibility is fully eliminated in this research.

The survey included multiple previously validated modules. First, in order to measure productivity and work satisfaction, the survey included the Health and Work Questionnaire (see Halpern et al., 2001; Shikiar et al., 2001, 2004). This easily-administered questionnaire allows for the assessment of various aspects of workplace productivity and its relationship to health. Following careful evaluation and confirmatory cluster analysis in Chapter 3, a revised version was validated, the Work Productivity and Stress Questionnaire (WPSQ), specifically fitting the working-from-home situation. This version outperforms the previous version from Shikiar et al. (2004) in consistency, reliability, and validation. Chapter 3 identifies five factors: productivity, productivity by others, peer relationships, nonwork satisfaction, and stress and irritability. For the remainder of the analysis, we use these newly constructed and validated productivity metrics.[46]

Shikiar et al. (2004) recommend validating its module with an alternative estimation of productivity. In addition to the questionnaire items, the survey therefore also included single-scale estimations of WFH productivity and satisfaction, such as self-reported productivity, satisfaction (with work in general, and with the WFH situation), and happiness. Additionally, participants stated their willingness to continue with WFH. All items were measured on a 10-point Likert scale, ranging from absolutely not (1) to completely (10).

The survey also included a short module estimating burnout propensity, comparable to Bloom et al (2015). Adopted from the Maslach burnout inventory (Maslach & Jackson, 1981), 6 questions were scored on a 7-point Likert scale, ranging from never (1) to always (7). These items, aimed at emotional exhaustion, include questions like "I feel empty at the end of a workday" and "I feel burned out.". Hence, a higher score on this metric indicates a higher burnout propensity. In addition to these six items, we added a 7-point Likert scale for sick days as well as break time during office hours. These additional items allow us to control for the previously established interaction between burnout and WFH, through the channels of increased

[46] For an extensive validation report, as well as a discussion on retrospective accuracy and validity, please see chapter 3.

autonomy (more breaks), or rather escalation into health issues (e.g. more sick days) compared to the pre-WFH situation.

To assess the conditions in the home office, we included two separate modules. First, we used the UC Berkeley center for building environment module, which has been developed and validated in order to assess the perceived indoor environmental quality (Zagreus et al., 2004). This survey has been extensively used in peer-reviewed research (e.g. Frontczak & Wargocki, 2011; Kim & de Dear, 2013; Palacios et al., 2020). The online assessment tool measures satisfaction on all relevant indoor environment factors such as indoor temperature, air quality, lighting, and noise. We also included relevant office hardware factors of the building environment module. These factors focus on the physical attributes in the office, pertaining to the desk, chair, screen, hardware, and Wi-Fi satisfaction. Note that, for convenience, we generally label all latter factors collectively as office hardware, although one of the variables of this factor is 'hardware'. The variable hardware specifically contains the pc, keyboard, and mouse and satisfaction. Throughout the paper, we generally refer to the factor and call it office hardware or office hardware satisfaction.[47] All factors are measured on a 7-point Likert scale, ranging from very dissatisfied (1) to very satisfied (7).

Next, we used a combination of work and office characteristics in the survey. These factors help to gain a complete picture of the WFH conditions and identify individual differences, potentially crucial factors influencing WFH outcomes. The office characteristics focused on the room (open versus closed), lighting (natural light versus no natural light), and ventilation (mechanic versus manual). All three factors are scored on a 3-point scale, with a neutral midpoint. Additionally, participants are asked to estimate the surface of their office (length and width in squared meters), and how often they ventilated their office (in percentage of time spent in the office). The work characteristics included the ability of the respondent to perform their work from home (1-10 scale), the company size (1-5, 5-15, 15-50, 50+ employees), length of the workweek in hours, and job category (e.g., governmental, non-governmental, self-employed, or on-call).

Finally, demographic information included age, gender, income, family size, and daily household situation. The daily household situation factors could support or hamper productivity when compared to the office situation. For instance, a dog walking break three times a day might increase productivity, whereas having to take care of a child might decrease productivity (van

[47] When discussing the variable, we explicitly specify that we are referring to the (sub) variable hardware.

der Lippe & Lippényi, 2020; Wells & Perrine, 2001). The house that participants reside in could interfere with the perceived WFH office characteristics. Matched on their 4-digit postcode, we added average urbanicity ('stedelijkheid'; STED), address-density ('omgevingsadressendichtheid'; OAD), and house value ('waarde van onroerende zaken'; WOZ) per participant.

4.4 Data

4.4.1 Descriptives – Demographics and Home Office

The survey was completed by 1,002 participants of which 58.1% are male and the mean age is 43.89 (SD=12,54). All participants had work that was at least partially executed from home, with 57.9% of the respondents exclusively working from home. 60.9% worked 36 hours a week or more, 84.8% had a fixed contract (self-employed and on-call workers represented 12.1% and 3.2%, respectively), with almost 70% at relatively large companies, with at least 50 employees. Table 4.1 shows further demographic characteristics. 54.6% of our sample enjoyed higher education (as compared to just over 40% for the Netherlands more broadly in 2019; Maslowski, R., 2020) and 53.6% earn more than the modal income in the Netherlands. These metrics support the notion that cognitively demanding (desk) jobs are more likely to be suitable to be performed from home (Etheridge et al., 2020). On average, our sample rated their jobs a 7.59 (out of 10; SD=2.39) on suitability to perform their work from home properly.

Table 4.1 Summary Statistics of Sample Demographics

		Mean (SD) or N (%)			Mean (SD) or N (%)
Demographics			Work Characteristics		
Age (years)		43.89 (12.54)	Income		
Gender (Female %)		420 (41.9%)		Modal wage (23-34k)	184 (18.4%)
Education Level				Minimum wage (less than 11,000)	23 (2.3%)
	Low	65 (6.5%)		Below modal (11-23k)	106 (10.6%)
	Middle	390 (38.9%)		1-2x modal (34-56k)	318 (31.7%)
	High			2x modal or more (56k)	219 (21.9%)
Family Characteristics				Don't know/ don't want to say	152 (15.2%)
Household Members		2.61 (1.21)	Company size (employees)		
Children Home during Office Hours				1-5	101 (10.1%)
	No Kids	482 (48.1%)		5-15	70 (7.0%)
	Always	33 (3.3%)		15-50	131 (13.1%)
	Sometimes	333 (33.2%)		50+	700 (69.9%)
	Never	154 (15.4%)	Work Sector		
Partner Home during Office Hours				Governmental	195 (19.5%)
	No Partner	240 (24.0%)		Non-governmental	654 (65.3%)
	Always	234 (23.4%)		Temp/ on-call worker	32 (3.2%)
	Sometimes	244 (24.4%)		Self-employed	121 (12.1%)
	Never	284 (28.3%)	Contract hours		
Pets				36+ hours	610 (60.9%)
	Dog	188 (18.8%)		20-35 hours	303 (30.2%)
	Cat	268 (26.7%)		12-19 hours	49 (4.9%)
				Less than 12 hours	40 (4.0%)
Home Office Characteristics			Work from Home Characteristics		
Home Office Floor Plan			Working from home currently		
	Open	377 (37.6%)		Exclusively from home	530 (57.9%)
	Average	139 (13.9%)		Partially from home	386 (42.1%)
	Closed	486 (48.5%)		Missing	86
Home Office Lighting				Work suitable to perform from home (0-10)	7.59 (2.39)
	Natural	828 (82.6%)	Work suitable to perform from home (0-10)		7.59 (2.39)
	Average	140 (14.0%)	House Characteristics		
	No Natural	34 (3.4%)	Real-estate value (WOZ; x €1.000))		274.64 (88.32)
Home Office Ventilation			Address-density (OAD; per 1 kilometer radius)		2118.10 (1749.55)
	Mechanic	135 (13.5%)	Urbanicity (STED; Categorical Address-density)		
	Manual	825 (82.3%)		Extremely High Urbanicity	199 (25.8%)
	None	42 (4.2%)		High Urbanicity	248 (32.2%)
Home Office Surface (m^2)		25.14 (17.40)		Average Urbanicity	154 (20.0%)
				Low Urbanicity	104 (13.5%)
				Non-Urban	66 (8.6%)
				Missing	231

Considering the home office, we find that they are relatively spacious (M=25.14 m², SD=17,40)[48] and predominantly illuminated by natural light (82,6%). These spaces vary between open (for instance, a studio; 37.6%) or closed (a dedicated room with a closed door; 48.5%), and

[48] Note that we use the estimated length and width of the office (in meters) to calculate the total surface in m². Extreme values (potentially mistakes) for either metric ultimately led to unrealistic outliers. As a result, we truncated the office surface from 2 to 100 m² (46 data points excluded).

almost all offices can be ventilated (82.3% manually, 13.5% mechanically). On average, participants exercise the option to ventilate 44% of their time at work in the work-from-home environment (SD=36.32).

We observe quite some variation in household conditions. On average, households consist of 2.60 people (SD=1.21), with a maximum family size of 9. Only 17.6% live in a single-person household (comparable to 18% in the Netherlands more broadly; CBS, 2022). We further specify the family situation during office hours: 47.8% have at least sometimes a partner around and 36.5% have children at home (of which roughly half has care responsibility, 16.1% of the total sample). Additionally, 18.8% of the respondents own a dog and 26.7% own a cat.

Table 4.2 Summary statistics and plots of the main variables of interest

	Mean	SD	Median	Min	Max	N	Boxplot	Histogram
Office Indoor Environment Satisfaction								
Temperature	5.13	1.28	5.00	1	7	1002		
Air Quality	5.41	1.12	6.00	1	7	1002		
Lighting	5.37	1.20	6.00	1	7	1002		
Noise	5.36	1.32	6.00	1	7	1002		
Office Hardware Satisfaction								
Desk	4.41	1.55	5.00	1	7	1002		
Chair	4.50	1.58	5.00	1	7	1002		
Screen	4.86	1.53	5.00	1	7	1002		
Hardware	5.19	1.32	5.00	1	7	1002		
WiFi	5.23	1.32	5.00	1	7	1002		
Ventilate During Office hours (%)	43.99	36.32	30.00	0	100	1002		
WPSQ Factor Scores								
Productivity	6.84	1.28	6.95	1.10	9.90	1002		
Productivity Others	7.55	1.24	7.75	1.00	10.00	957		
Stress and Irritability	3.82	1.63	3.64	1.00	9.21	1002		
Peer Relations	6.65	1.59	6.67	1.00	10.00	968		
Non-work Satisfaction	5.99	1.59	6.33	1.00	10.00	1002		
Single-item Scales Scores								
Satisfaction Work Situation	6.82	1.90	7.00	1	10	1002		
Happy with Work Situation	6.76	1.96	7.00	1	10	1002		
Self-Reported Productivity	7.16	1.72	7.00	1	10	1002		
Burnout Propensity Score								
Burnout Metric	2.87	1.25	2.67	1.00	7.00	1002		

4.4.2 Non-Parametric Comparisons between Home Office and the Work Office

We measure productivity both through a factor based on a set of questions (WPSQ factor), as well as single questions on productivity indicators, as suggested by Shikiar et al. (2004). Table 4.2 shows the general scoring on the main variables of interest, as well as the distribution by boxplot and histogram, for the home office situation. Table 4.3 compares the home office scores with the work office scores by applying nonparametric Wilcoxon signed-rank tests on paired samples' median differences. The average WPSQ factor productivity score at home is

6.84 out of 10 (SD=1.28 with a maximum of 9.90). The participants expect that others score their productivity significantly lower (M=7.55, SD=1.24; p=.01). Stress and irritability are on average rated at 3.82 (SD=1.63). Non-work satisfaction and peer relations score around 6 (SD=1,59) and 6.65 (SD=1,59), respectively. Compared to the work office, the WPSQ factor productivity scores higher at work (p<.001), whereas self-reported productivity does not differ (p >.06). The overall trend for the other WPSQ factors (excluding Stress) shows a higher score for the work office.[49]

Table 4.3 Satisfaction and Productivity Non-Parametric Comparison Tests: Home Office versus Work Office

	Home Office (N=1002)	Work Office (N=1002)	p-value
Office Indoor Environment Satisfaction *(scale: 1-7)*			
Temperature	5.13 (1.28)	4.59 (1.24)	0.00***
Air Quality	5.41 (1.12)	4.61 (1.27)	0.00***
Lighting	5.37 (1.20)	5.07 (1.30)	0.00***
Noise	5.36 (1.32)	4.63 (1.34)	0.00***
Office Hardware Satisfaction *(scale: 1-7)*			
Desk	4.41 (1.55)	5.46 (1.12)	0.00***
Chair	4.50 (1.58)	5.37 (1.14)	0.00***
Screen	4.86 (1.53)	5.52 (1.09)	0.00***
Hardware	5.19 (1.32)	5.44 (1.07)	0.00**
WiFi	5.23 (1.32)	5.52 (1.15)	0.00***
	Home Office (N=1002)	Work Office (N=1002)	p-value
WPSQ Factor Scores *(scale: 1-10)*			
Productivity	6.84 (1.28)	7.11 (0.93)	0.00***
Productivity by Others	7.55 (1.24)	7.78 (1.04)	0.01*
Stress and Irritability	3.82 (1.63)	3.95 (1.55)	1.00
Peer Relations	6.65 (1.59)	7.41 (1.23)	0.00***
Non-work Satisfaction	5.99 (1.59)	7.59 (1.06)	0.00***
Single-Item Scale Scores *(scale: 1-10)*			
Satisfaction Work Situation	6.82 (1.90)	7.22 (1.62)	0.00***
Happy with Work Situation	6.76 (1.96)	7.30 (1.56)	0.00***
Self-Reported Productivity	7.16 (1.72)	7.47 (1.35)	0.06
Burnout Propensity *(scale: 1-7)*			
Burnout Metric	2.87 (1.25)	2.64 (1.10)	0.00***

Note. P-values result from Kruskal-Willis rank-sum tests. After Bonferroni correction: *p<0.05, **p<0.01, ***p<0.001. Work office WPSQ, single-itemscale, and burnout scores are collected during a pretesting phase. Not all participants participated in the pretest, therefore a fraction of the work office scores could not be matched to their respective home office scores.

[49] Note that these differences are, although significant (due to the large sample size), relatively small to the extent that the economical relevance is limited. As such, we are cautions with generalizing the differences as significant main effects.

The single-question estimations of productivity and satisfaction show a slightly higher, yet similar, trend. Self-reported productivity, satisfaction, and happiness at home are all around 7 out of 10 on average (M=7,16, SD=1,72; M=6,82, SD=1.90; M=6,76, SD=1,96, respectively). Appendix Table 4.6 shows that the WPSQ factor productivity estimator is strongly correlated with its single-question counterpart (r=.73, p<.0001) suggesting a high convergent validity. Throughout the remainder of the paper, we solely refer to the WPSQ productivity factor when we discuss productivity scores.

The average burnout score suggests that the majority of our sample shows limited signs of burnout during the home office (on a 7-point scale; M=2,87, SD=1,25). This score, however, does not deviate far from similar reports of a larger Dutch sample, which uses the same measurement (Peijen et al., 2022). Yet, relative to home, the work office performs better: at home, burnout propensity is significantly higher compared to the work office (p<.01). The burnout score is highly correlated with the WPSQ factor "stress and irritability" (Appendix Table 4.6; r=.71, p<.0001). The sample showed no significant difference in either sick days or break time between the work office and the home office.[50]

Importantly, participants score 6.25 out of 10 (SD=2,93) on their willingness to continue with homework in the current situation (see Table 4.2). Although not low in absolute terms, this score seems relatively low in contrast to the optimistic sentiment stated by many polls following the COVID-19 WFH obligations. In addition to the 97% of 2,300 state to at least partially switch to WFH (Griffis, 2021), under over 3,500 U.S. workers, 68% would choose to work from home over the office in general, with 61% even prepared to accept a pay cut to maintain that WFH situation (Korolevich, Sara, 2021). This 68% preference for WFH over the office signals that our average willingness to continue of 6.25 is an accurate estimation. In sum, the burnout score and willingness to continue to appear to accurately estimate the population, which help us generalize our results to the general population.

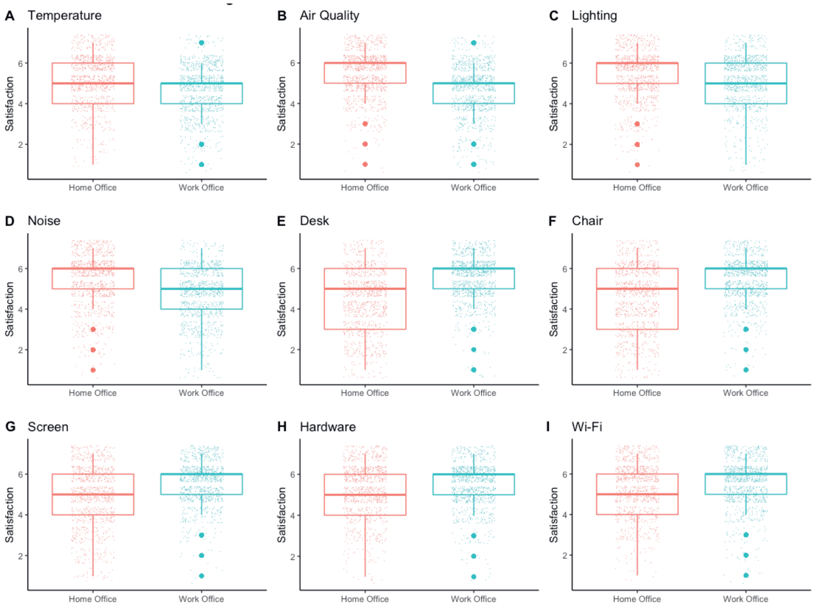
Nevertheless, the large variance in our sample scoring at indicates that the score is not uniform, and this variation needs further explaining. Therefore, we will include the willingness to continue working from home in our final analysis. In addition, the self-reported satisfaction and happiness with the working situation at home versus working from the work office, displayed in Table 4.3, also shows that working from the office is still preferred relatively high(er) compared to working from home.[51] However, it must be noted that at those scores are

[50] These metrics are not therefore not further included in this paper.
[51] Note that the same limitation mentioned in footnote 4 apply to these score differences.

influenced by the fact that at that time, working from home was mandatory and not born out of preference.

The office hardware and office indoor environment factors are investigated for both the home office and work office. Figure 4.2 shows the distribution plots of both office indoor environment (A-D) and office hardware (E-I) scores. WFH increased the satisfaction of all office indoor environment factors: Temperature (A), Air Quality (B), Lighting (C), and Noise (D) all scored higher as compared to the work environment (mean scores range between 5.37 and 5.13 for the home office, compared to 5.07 and 4.59 for the work office; on a 7-point Likert scale). For office hardware, we observe the opposite trend: overall office hardware satisfaction at the work office is higher. The satisfaction for the desk (E), chair (F), screen (G), hardware (H), and Wi-Fi (I) range between 5.23 and 4.41 at home, whereas the work office hardware satisfaction levels range between 5.52 and 5.37. Table 4.3 shows that all differences are statistically significant, using the non-parametric Wilcoxon rank sum test and Bonferroni multiple comparison corrections. These observations support the notion that at home, optimizing ergonomics (e.g office hardware factors) remains challenging (Davis et al., 2020) while increased individual control over office indoor environment is preferred (Chang & Kajackaite, 2019).

Figure 4.2 Office Hardware and Indoor Environment Satisfaction boxplot for Home Office and Work Office



It is important to confirm that participants are considering and rating the home office as distinctly different from their work office. We correlate each variable's score at home and at the office. As shown in Appendix Table 4.6, scores correlate moderately with different variables within the same environment (home office or work office), but correlate low with the same variables in the other environment. In other words, the same environment seems to influence the scores more than which variable is scored over both environments. For instance, the correlation between temperature and noise at home is r=0.41, which is considered a moderately strong correlation (Schober et al., 2018). Comparatively, the correlation between the temperature at the work office and the home office is negligible (r=0.06). This observation supports the divergent validity of the survey responses. Confirming that the satisfaction scores are based on different experiences at home compared to work confirms the relevance of home office satisfaction to predicting home productivity and burnout propensity.

In conclusion, our comparison indicates that our sample seems to fare better at the work office: their productivity, non-work satisfaction, and peer relations are all higher compared to the home office. Burnout propensity increases during home office. Generally, happiness and satisfaction with the work situation are higher for the work office. The assessment of the physical office shows that the office hardware and office indoor environment display contrasting satisfaction trends: hardware is preferred at the work office, whereas the indoor environment is preferred at home. In the next section, we will investigate how these results relate to each other. Whilst controlling for an array of relevant factors, we examine whether physical office satisfaction affects productivity-related scores.

4.5 Methods

Whether differences in home-work conditions lead to productivity depends on many factors. We use multiple linear regression to formally assess the relationship between home office hardware and home office indoor environment satisfaction on productivity and burnout propensity using the following models:

$$y_i = \alpha_0 + \alpha_1 \text{hardware satisfaction}_i + \alpha_2 GC_i + \varepsilon_i \tag{1a}$$

$$y_i = \beta_0 + \beta_1 \text{indoor environment satisfaction}_i + \beta_2 GC_i + \varepsilon_i \tag{1b}$$

where y_i is the predicted value of either productivity or burnout propensity for each participant *i*. Model 1a isolates the effect of home office hardware ($\text{hardware satisfaction}_i$), whereas model 1b isolates the effect of the home office environment ($\text{indoor environment satisfaction}_i$) on our dependent variables. Both models include a set of carefully selected general controls (GC_i) that could otherwise confound the estimators. Specifically, these include demographic characteristics (gender and age), job characteristics (company size, job suitable for working from home, type of work, income, and working hours), and household characteristics (household size, children at home, partner at home, and pets)

Model 2 shows the combined model including both the effect of home office hardware and home office indoor environment on our dependent variable y_i .

$$y_i = \delta_0 + \delta_1 \text{hardware satisfaction}_i + \delta_2 \text{indoor environment satisfaction}_i + \delta_3 GC_i + \delta_4 OC_i + \varepsilon_i \tag{2}$$

This model also stepwise adds physical characteristics of the home office controls (OC_i), including lighting (natural versus no natural), means of ventilation (none, manual, or ventilation), and the room plan (open versus closed).[52] Running model 2 with and without home office controls, we estimate four models in total for both productivity and burnout propensity.

For all models, we standardized all continuous variables since they are originally measured on different Likert scales, to simplify the interpretation of the coefficients.[53] As a result, the coefficients are z-scores and must be interpreted such that each coefficient indicates the change in de dependent variable for each standard deviation increase of the independent variable. Moreover, Figure 4.2 shows that some office hardware factors display comparable trends in scoring. By further inspection, correlation Table 4.7 in the Appendix shows that both desk and chair, as well as screen and the hardware factor, have an internal correlation (*r*) exceeding 0.70. Since the correlations between these variables are intuitively not surprising, they are comfortably specified as a combined variable model. Thus, for any further analysis, the scores on these two pairs are combined and averaged per participant.

4.6 Results

4.6.1 Explaining Productivity and Burnout in the Home Office

Table 4.4 shows the estimated standardized coefficients and standard errors of the home office hardware and home office indoor environment variables on the productivity factor. The results in Table 4.4 show that all office hardware variables at home are positively associated with productivity, such that increased satisfaction with each office hardware variable is associated with an increase in productivity when working from home (coefficients ranging from 0.18 to 0.15; SD = .03 to .05). For example, model 1 shows that for each standard deviation increase of screen & hardware satisfaction, the productivity score increases by 0.18 standard deviations (SD=.05). Using Wi-Fi satisfaction as an example, a 1.32 increase on a 0-7 satisfaction scale translates to a 0.23 increase on a 0-10 productivity scale. This effect is relatively strong and comparable in strength to moving from sometimes having children at home during working hours to having no children at all (see Appendix Table 4.8). Home office indoor environment

[52] In the Appendix, an additional model is shown, for which we match our participants on postcode level to mean house characteristics. These included average urbanicity, density, and WOZ value. None of these variables are significant for either Productivity nor Burnout, nor do they influence the coefficients in a noteworthy matter.

[53] Coefficients are standardized unless specifically mentioned otherwise

variables show a similar pattern: without exception, all variables are associated with increased productivity (coefficients ranging from 0.21 to 0.08; SD = .04). Combining both home office hardware and indoor environment variables in model 3 decreases the coefficients intensity for some variables in the productivity model. Incidentally, air quality loses significance. Additionally adding controls in model 4 hardly affects the model further. All office hardware variables are relevant predictors of productivity, as well as temperature and noise satisfaction (indoor environment).

Table 4.4 Regressions of office hardware and indoor environment satisfaction on productivity

	Productivity			
	(1)	(2)	(3)	(4)
Desk & Chair	.15 (.04)***		.09 (.04)**	.09 (.04)**
Screen & Hardware	.18 (.05)***		.11 (.05)**	.11 (.05)**
WiFi	.18 (.03)***		.10 (.04)***	.10 (.04)***
Temperature		.14 (.04)***	.09 (.05)**	10 (.05)**
Air Quality		.08 (.04)*	.03 (.04)	.04 (.04)
Lighting		.11 (.04)***	.07 (.04)*	.06 (.04)
Noise		.21 (.04)***	.16 (.04)***	.16 (.04)***
General Controls	Yes	Yes	Yes	Yes
Home Office Controls	No	No	No	Yes
Observations	1,002	1,002	1,002	956
R^2	.25	.27	.30	.30
Adjusted R^2	.23	.25	.28	.27
Residual Std. Error	.88 (df = 972)	.87 (df = 971)	.85 (df = 968)	.85 (df = 915)
F Statistic	11.41*** (df = 29; 972)	12.16*** (df = 30; 971)	12.79*** (df = 33; 968)	10.01*** (df = 40; 915)

Note. *p<0.1, **p<0.05, ***p<0.01. For the full model including all controls, see Appendix Table 4.8

Table 4.5 shows the coefficients for home office hardware and home office indoor environment satisfaction on burnout propensity. Again, models 1 and 2, estimate the home office hardware and home office indoor environment coefficients separately, model 3 combines both, and model 4 adds physical characteristics of the home office as controls. For the burnout models, the association is negative, meaning that an increase in satisfaction on either variable's satisfaction is associated with a decrease in the individual level of feeling burnout. For instance, model 1 shows that each standard deviation decrease in WI-FI satisfaction increases the burnout score by 0.12 standard deviation (SD=.04). Similarly to productivity, combining both satisfaction predictors in one model (model 3), decreases the significance of some variables. The most robust predictors of burnout propensity are desk & chair and Wi-Fi satisfaction (home office hardware), as well as air and noise satisfaction (home office indoor environment).

Table 4.5 Regressions of office hardware and indoor environment satisfaction on burnout propensity

	Burnout Propensity			
	(1)	(2)	(3)	(4)
Desk & Chair	−.14 (.04)***		−.10 (.04)**	−.11 (.05)**
Screen & Hardware	− .09 (.05)*		−.04 (.05)	−.03 (.05)
WiFi	−.12 (.04)***		−.07 (.04)*	−.07 (.04)*
Temperature		−.07 (.04)*	−.04 (.04)	−.06 (.04)
Air Quality		−.11 (.04)**	−.08 (.04)*	−.09 (.04)**
Lighting		−.05 (.04)	−.02 (.04)	.01 (.04)
Noise		−.13 (.04)***	−.10 (.04)***	−.09 (.04)**
General Controls	Yes	Yes	Yes	Yes
Home Office Controls	No	No	No	Yes
Observations	1,002	1,002	1,002	956
R^2	.18	.19	.21	.21
Adjusted R^2	.16	.17	.18	.17
Residual Std. Error	.92 (df = 972)	.91 (df = 971)	.91 (df = 968)	.90 (df = 915)
F Statistic	7.60*** (df = 29; 972)	7.71*** (df = 30; 971)	7.70*** (df = 33; 968)	6.06*** (df = 40; 915)

Note. **p*<0.1, ***p*<0.05, ****p*<0.01. For the full model including all controls, see Appendix Table 4.9.

Comparing both tables shows that, on average, office hardware and indoor environment coefficients and significance levels are generally higher in the productivity models. For example, noise satisfaction is meaningful for both productivity as well as burnout propensity, yet the coefficient is roughly 50% higher for productivity for all models (0.21 to 0.16 versus -0.13 to -0.09, for productivity and burnout, respectively).

Factors other than hardware and indoor environment (e.g. household characteristics), used in the regressions of Table 4.4 and Table 4.5 as control variables, are consistently associated with productivity and burnout propensity. These factors give insight into the baseline propensity for WFH productivity and burnout levels, beyond hardware and indoor environment. Both models 3 of Appendix Table 4.8 and

Table 4.9 show that the degree to which work can be performed from home does not add predictive value to our model. Contrasting Etheridge et al. (2020), yet in line with Awada et al. (2021), we find that women tend to report higher levels of productivity (δ = 0.15, SD= .07). Not living alone, i.e. having a larger household, decreases burnout score and increases productivity (δ = -0.10, SD=.04; δ = 0.11, SD= .04, respectively). Interestingly, and only significant for productivity, having a partner who is not (or only sometimes) home during office hours increases productivity (δ = 0.14 - 0.15, SD=.08) compared to the baseline of having no partner at all. In that sense, having a partner is good for productivity, as long as they are not constantly present during office hours. For children, a more predictable, strong, and linear pattern emerges: burnout propensity increases and productivity decreases when potential children spend more time at home during office hours. Interestingly, having a dog increases the burnout score significantly (δ = 0.17, SD=.08).

Finally, previous research indicated that during the pandemic, young workers seem to appreciate WFH more, and opted for it more often compared to older workers (Brynjolfsson et al., 2020). These results, however, are not stable per se. Awada et al. (2021) show opposite results, where both women and older workers reported being more productive during WFH. At first glance, our results support the Awada et al. findings to the extent that for each year older, reported productivity increases with 0.01 standard deviation (SD = .003)[54]. This effect on productivity translates such that moving from 20 years old to 40 years old increases WFH productivity score with roughly 0.25 out of 10. In strength, this example's effect is double as strong as the gender effect or the individual office conditions effect on productivity. However, we find a weak yet similar trend for reported productivity at the office, which does not provide insight on the preference of WFH over the work office.[55] Looking at the willingness to continue with WFH, we finally find older responders reporting a stronger willingness to continue to work from home (0.01 standard deviation increase per year of age, SD = .003; see Appendix

[54] Note that age is not standardized in these models.

[55] This table is not reported in this paper. With productivity at work as dependent variable in a linear regression similar to Appendix Table 4.8, we find a positive coefficient of .001 (SD=.003, p=.0445) for model 3.

Table 4.10). Therefore conclusively, our results suggest that older workers not only report to be more productive at home and at the office than younger workers, but also seem to have an overall higher willingness to continue to work from home.

4.6.2 Integrative model

Home office conditions, including hardware factors such as Wi-Fi and desk setup, as well as indoor environmental quality, are significantly associated with productivity and burnout propensity. While some of these input factors are fixed or dependent on capital expenditures, indoor environmental conditions depend to a large extent on human behavior. We, therefore, aim to understand whether there are causal indicators for satisfaction with hardware and indoor environment. Importantly, we measure the behavior of respondents working from home: whether the home office is ventilated, both at the extensive and intensive margin.

We implement a mediation analysis in order to understand how the home office environment influences productivity. First, we test whether ventilation, by means of improving the objective air-quality, leads to higher self-reported productivity in line with recent research{Citation}. Then, we assess how physically improving the indoor environment relates to the indoor satisfaction effect on productivity. Improving the indoor air quality through ventilation should improve the perceived indoor environment satisfaction, thus at least partially mediating ventilation's effect on productivity. We also asses the mediating power of unrelated satisfaction variables such as hardware. If self-report is accurate, these should be unrelated to ventilation. However, if these scores also mediate the effect of ventilation, objective indoor improvements could increase general satisfaction, and specific satisfaction scores could be less reliable in representing the underlying elements.

This analysis assesses the impact of the physical environment on productivity, mediated by hardware and indoor environment satisfaction factors. For the analysis, we construct two latent variables, 'Office Hardware' and 'Office Indoor Environment', that each consist of all individual hardware and indoor environment satisfaction variables (see Appendix Figure 4.5 for the loadings per latent variable). Confirmatory factor analysis shows that the 'Office Hardware' and 'Office Indoor Environment' items loadings are meaningful per latent factor.[56] Further reliability calculations confirm the factor's consistency, with both factors

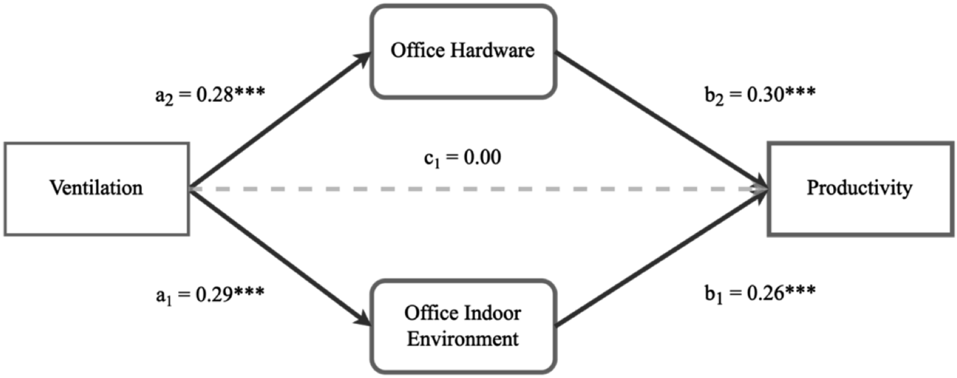
[56] The factors are loaded by the marker variable identification approach. By doing so, the estimators of the latent variables on the dependent variable are fixed on the original 7-point satisfaction. In other words, the estimators indicate the effect per point estimate increase on a 7-point scale identical to the scales of the underlying variables.

showing a Cronbach alfa above 0.8 (a=0.80 and a=0.85, for 'Office Hardware' and 'Office Indoor Environment', respectively). Following model specification analysis, we declare covariance between the latent variables 'Office Hardware' and 'Office Indoor Environment', and indicator items desk and chair as well as screen and hardware. Since the correlations between these variables are intuitively not surprising, they are comfortably specified in a saturated model. This saturated model, containing additional parameters estimating those correlations, indeed fits the data better than the restricted model with these correlations fixed to zero (chi-squared difference = 568, DF difference = 3; $p<.000$).[57] Further model fit tests, confirm that our saturated model fits the data well (CFI/TLI > .95, RMSEA close to .05, and SRMR < .05).

As expected, the latent variables 'Office Hardware' and 'Office Indoor Environment' have a strong and distinct direct effect on WFH productivity, as can be seen in Figure 4.3. For both factors, a standard deviation increase is associated with around a 0.3 standard deviation increase in productivity. The percentage of ventilation is significantly associated with both increased hardware and indoor environment satisfaction. Each standard deviation increase in ventilation of the office increases satisfaction with .29 and .27 points, respectively. Ventilation no longer shows a direct association with productivity, which is not captured by its relation to hardware or indoor environment satisfaction ($p=.88$). Hence, the effect of ventilation on productivity is fully mediated by hardware or indoor environment satisfaction. Both indirect unstandardized parameters via the latent variables are estimated at 0.002, with a total estimated effect of ventilation on productivity of 0.004. Thus, moving from 0% to 100% ventilation of the office increases productivity with .4 on the 10-point scale by increasing hardware or indoor environment satisfaction. Considering that the average productivity score is 6,11 (SD=1,06), the magnitude of this effect is not trivial. This effect equates to 8.18% of the mean and 47% of the standard deviation of the productivity variation in our sample.

[57] Note that we do not combine the pairs desk & chair and screen & hardware pre-analysis in contrast to the multivariate regression, but enter them individually whilst declaring covariance in the SEM model. Doing so increases the Cronbach alpha of both models with 0.05 and improves the overall model fit.

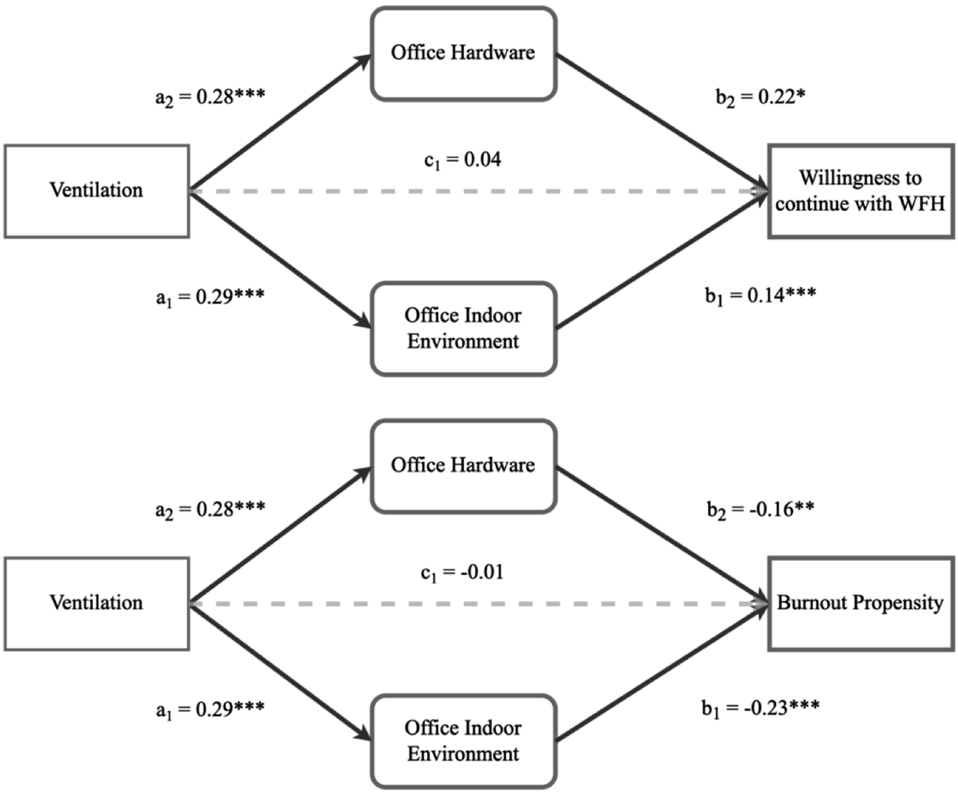
Figure 4.3 Structural Equation Model Graphs on Productivity



Note. **p*<0.05, ***p*<0.01, ****p*<0.001

Replacing productivity with burnout propensity or willingness to continue WFH in the model shows the same mediation effect. Both models, are well fitted (both show CFI/TLI > .95, RMSEA close to .05, and SRMR < .05), and for both models, the association runs fully through the latent variables (see Figure 4.4). The total estimated effect of ventilation on burnout propensity is -0.004, with comparable mediation through both factors. Moving from 0% to 100% ventilation of the office decreases burnout propensity with .4 on the 7-point scale. For the willingness to continue with WFH, the significance and strength of association are stronger for hardware compared to the indoor environment (a=0.003, p=0.016; a=0.005, p<.000, respectively). Hence, moving from 0% to 100% ventilation of the office increases the willingness to continue WFH propensity with 1.2 on the 10-point scale.

Figure 4.4 Structural Equation Model Graphs on Burnout Propensity and Willingness to Continue WFH



Note. $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$

4.7 Conclusion

The success of working from home, and the likelihood of its survival after the pandemic, is dependent on sustained productivity in the home office environment. This study investigates the effect that home office hardware and environment satisfaction have on productivity and burnout propensity. Through a survey design, we gather data on the home offices, home office satisfaction levels, and productivity from 1,002 participants who have been working from home. First, we compare working from home with working from the office. The self-reported productivity is lower at home compared to previously at the office of those that worked from home during the COVID-19 pandemic. Although WFH reports are predominantly positive, a growing body of research has shown the limits and downsides of WFH. Reports that included actual output assessments suggest less or no productivity improvement (Adrjan et al., 2021; Aksoy et al., 2022; Barrero et al., 2021; Bloom et al., 2022; Etheridge et al., 2020; Gibbs et al., 2021; Morikawa, 2022). This is in contrary to earlier findings based on self-report, but consistent with multiple non-self-reported outcome analysis (Etheridge et al., 2020; Gibbs et al., 2021; Morikawa, 2022). When looking at the physical office, we find that the indoor environmental satisfaction appears higher at home, whereas physical hardware satisfaction like desks and chairs are preferred at the office. From these results, we conclude that optimizing ergonomics at home remains challenging (Davis et al., 2020) while individually being in control of the indoor environment at home is preferred (Chang & Kajackaite, 2019). Additionally, we find a relatively low score for the willingness to continue working from home, in contradiction to many recent reports, which supports a deeper investigation into factors facilitating successful WFH (Griffis, 2021; Korolevich, Sara, 2021).

Second, the results show a consistent association between both home office hardware as well as indoor environment satisfaction and productivity. Higher satisfaction in both these domains is associated with higher WFH productivity and lower burnout propensity. The vast majority (with the exception of air quality) of all indoor environment and hardware factors included in this paper are associated with increased productivity. Given our results of ventilation, it is particularly striking that precisely air quality is the only self-reported factor seemingly not influencing WFH success. For burnout propensity, we find the most robust predictors to be desk & chair (hardware), as well as noise satisfaction (indoor environment). Additionally, we find women and larger households to be more productive at home, having children to decrease productivity and increase burnout propensity, and partners to increase productivity, but only when they are not always around during office hours. Finally, where the correlation between age and WFH has been inconsistent in the past (Awada et al., 2021; Brynjolfsson et al., 2020), we find older workers report being more productive, having a low burnout propensity, and stating to be more willing to continue to work from home compared to younger workers.

To show the influence real behavior could have on WFH success, we investigate the effect of ventilation on productivity. By means of mediation analysis, we confirm that the amount of time a home office is ventilated not only directly increases satisfaction but indirectly increases the productivity score as well, with a total estimated effect of 0.005 points per percentage-point increase. Practically, this means that changing from not ventilating at all to ventilation all the time (moving from 0% to 100%), will indirectly increase productivity with .5

points on the 10-point productivity scale. The magnitude of this estimate on productivity is comparable to moving from no children to always having children at home during working hours (.7-point decrease of productivity). Given the same example, moving from 0% to 100% ventilating time will decrease the burnout propensity by .4 points on a 7-point scale, and increase the willingness to continue with WFH with 1.2 points on a 10-point scale. Hence, we find that ventilating your home office is a crucial underlying factor explaining overall satisfaction and is indirectly associated with increased productivity, increased willingness to work from home, and decreased burnout propensity.

The results of this paper imply that the success and willingness to continue with work from home are at least partially dependent on home office quality and satisfaction. We show that both the hardware and indoor environment satisfaction as well as the behavior influencing the office environment quality will contribute to higher productivity, lower burnout propensity, and higher willingness to continue working from home. Incidentally, we also provide a strong case to emphasize actual measurement over self-reported satisfaction measurement. This is not only shown by the fact that unrelated satisfaction scores are influenced by better ventilation, but also by the fact that self-reported air quality satisfaction, closest related to ventilation, did not predict productivity. Since the general working from home evaluation, as well as indoor satisfaction inventories, are heavily reliant on self-reported scores, this conclusion is not trivial. Consequently, satisfaction with unrelated aspects of the office, as well as WFH success, can be influenced (and thus improved) by seemingly unrelated actions such as increasing office ventilation. In sum, our results underline the effect of a holistic perspective on working from home productivity. A healthy and objectively measured physical climate is a key aspect of the success and widely proposed bright future of working from home (Adrjan et al., 2021; Aksoy et al., 2022; Hansen et al., 2022).

Our results have some limitations. First, our results follow an unusual and unique situation that has not been experienced before. This can have an effect on all data collected during the COVID-19 pandemic, yet our data is especially sensitive to the sentiment of our sample. We have attempted to control for an extensive set of factors that might influence WFH and productivity in general, however the pandemic will undoubtedly have had an immense effect on daily life. We cannot exclude the possibility that unobserved factors, or even unobserved general sentiment (mood-as-information theory; O'Donnell et al., 2020), influences the self-reported scores. Additionally, the experienced ability to visit an office when needed alone could also influence the perceived productivity. As such, the context of the pandemic could not only influence the general sentiment, but also the operational experience of WFH which could be subject to change when restrictions are lifted.

Second, we report on differences between the current situation and before COVID-19 (at the work office). To do so, we did not ask our participants at that time, but rather asked them recollect. Unfortunately, recollection itself is less accurate than asking the current situation (Ingram et al., 2012). The current situation could even influence the recollected score, as it serves as a reference point (Joordens & Hockley, 2000). Either in contrast or regress, it influences the absolute score. As this is already normally the case, we ask our participants to recollect pre-COVID-19. This means that not only the working from home sentiment can pass over to the recollection, but the general sentiment as a whole. The mere fact that WFH is mandatory could put the productivity previously performed at work in a more generous daylight that it truly was, as well as life in general. Taken together, although the pandemic also provides us with a unique naturally occurring experimental setting, our data quality would have improved if we would have foreseen the pandemic, and pretested our subject before the outbreak.

In the same line, although we use an elaborate and implicit measure for productivity and burnout, we do not measure objective productivity. Self-reported metrics suffer from demand effect and accuracy biases which might hamper the accuracy of our result in the general population (Chapter 3). We have attempted to at least partially alleviate this concern by using an extensive validated questionnaire, yet caution to generalize to actual behavior is warranted.

4.8 Appendix

Table 4.6 Correlation Table of different measurements of Productivity and Stress

	WPSQ Productivity	Self-Reported Productivity	WPSQ Stress & Irritability Factor
Self-Reported Productivity (Single-item Scale)	.73		
WPSQ Stress & Irritability Factor	-.55	-.37	
Burnout Propensity (MBI Scale)	-.50	-.31	.71

Table 4.7 Correlation Table of home office hardware and indoor environment factors' Satisfaction

	Desk Home	Chair Home	Screen Home	Hardware Home	Wi-Fi Home	Temperature Home	Air Quality Home	Lighting Home	Noise Home
Panel 1: Home Satisfaction									
Chair Home	.70								
Screen Home	.57	.58							
Hardware Home	.50	.52	.72						
Wi-Fi Home	.36	.39	.46	.55					
Temperature Home	.36	.33	.36	.39	.43				
Air Quality Home	.35	.38	.39	.45	.42	.60			
Lighting Home	.36	.32	.35	.37	.37	.53	.60		
Noise Home	.35	.36	.37	.40	.37	.41	.48	.41	
Panel 2: Work Satisfaction									
Desk Work	.09	.21	.18	.24	.21	.14	.23	.14	.21
Chair Work	.12	.17	.19	.25	.22	.19	.25	.18	.20
Screen Work	.07	.17	.17	.25	.27	.19	.27	.19	.22
Hardware Work	.14	.22	.26	.37	.32	.21	.29	.19	.26
Wi-Fi Work	.17	.24	.23	.33	.28	.18	.24	.18	.25
Temperature Work	.10	.14	.11	.15	.13	.06	.14	.12	.05
Air Quality Work	.13	.15	.15	.17	.13	.13	.17	.13	.08
Lighting Work	.10	.17	.20	.19	.15	.19	.24	.17	.16
Noise Work	.06	.08	.12	.12	.14	.15	.18	.13	.04

Table 4.8 Complete Regression Table of Productivity

	Dependent variable: Productivity				
	(1)	(2)	(3)	(4)	(5)
Desk & Chair	.15 (.04)***		.09 (.04)**	.09 (.04)**	.12 (.05)**
Screen & Hardware	.18 (.05)***		.11 (.05)**	.11 (.05)**	.12 (.05)**
WiFi	.18 (.03)***		.10 (.04)***	.10 (.04)***	.07 (.04)
Temperature		.14 (.04)***	.09 (.05)**	.10 (.05)**	.10 (.06)**
Air Quality		.08 (.04)*	.03 (.04)	.04 (.04)	.08 (.05)
Lighting		.11 (.04)***	.07 (.04)*	.06 (.04)	.04 (.05)
Noise		.21 (.04)***	.16 (.04)***	.16 (.04)***	.13 (.04)***
Age (years)	.01 (.003)***	.01 (.003)***	.01 (.003)***	.01 (.003)***	.01 (.003)***
Income (Baseline: Modal)					
Mininum wage (less than 11,000)	-.68 (.24)***	-.56 (.24)***	-.58 (.23)***	-.58 (.24)***	-.46 (.28)**
below modal (11-23k)	-.10 (.10)	-.03 (.10)	-.04 (.10)	-.06 (.10)	-.05 (.11)
1-2x modal (34-56k)	-.12 (.08)	-.10 (.08)	-.10 (.08)	-.09 (.08)	-.08 (.09)
2x modal or more (56k)	-.14 (.09)	-.11 (.09)	-.12 (.09)	-.12 (.09)	-.02 (.10)
don't know/ don't want to say	.01 (.09)	.04 (.09)	.01 (.09)	-.01 (.09)	-.01 (.10)
Female	.21 (.07)***	.10 (.07)	.15 (.07)**	.14 (.07)**	.15 (.08)**
Household Members	.10 (.04)**	.12 (.04)***	.11 (.04)***	.09 (.04)**	.07 (.05)
Children Home during Office Hours (baseline: no children)					
Always	-.70 (.19)***	-.65 (.18)***	-.63 (.18)***	-.63 (.18)***	-.74 (.21)***
Sometimes	-.22 (.10)**	-.23 (.10)**	-.22 (.09)**	-.17 (.09)*	-.17 (.10)
Never	-.07 (.10)	-.07 (.10)	-.07 (.10)	-.06 (.10)	-.07 (.11)
Partner Home during Office Hours (baseline: no Partner)					
Always	.06 (.09)	.06 (.09)	.06 (.09)	.06 (.09)	.02 (.10)
Sometimes	.17 (.09)*	.11 (.09)	.15 (.08)*	.13 (.09)	.13 (.10)
Never	.20 (.09)**	.11 (.09)	.14 (.08)*	.15 (.09)*	.10 (.10)
Pet (Baseline: No pets)					
Dog	.10 (.07)	.09 (.06)	.07 (.06)	.07 (.07)	.08 (.08)
Cat	-.02 (.06)	-.01 (.07)	-.02 (.06)	-.04 (.06)	-.01 (.08)
Company size (Baseline: 0-5)					
5-15	.01 (.17)	-.04 (.16)	-.05 (.16)	-.09 (.17)	-.16 (.20)
15-50	.10 (.15)	.02 (.14)	.02 (.14)	-.02 (.15)	-.12 (.18)
50+	-.003 (.14)	-.03 (.14)	-.04 (.14)	-.08 (.14)	-.14 (.18)
Work Sector (Baseline: Governmental)					
Yes, non-governmental	-.03 (.07)	.02 (.07)	.0005 (.07)	-.01 (.07)	-.01 (.08)
Yes, temp/ on-call worker	.03 (.17)	.14 (.17)	.09 (.16)	.09 (.17)	.05 (.17)
Yes, self-employed	-.04 (.15)	.01 (.14)	-.03 (.14)	-.10 (.15)	-.14 (.17)
Contract hours (Baseline: Full time (36+)					
20-35 hours	-.02 (.07)	-.03 (.07)	-.03 (.07)	-.01 (.07)	.004 (.08)
12-19 hours	.20 (.13)	.16 (.13)	.18 (.13)	.20 (.13)	.19 (.14)
less than 12 hours	-.05 (.13)	-.18 (.12)	-.12 (.12)	-.11 (.13)	-.01 (.13)
Work suitable to perform from home	.05 (.03)	.07 (.03)**	.04 (.03)	.04 (.03)	.06 (.04)*
Home Office Floor plan (Baseline: Average)					
Open				.02 (.10)	.05 (.11)
Closed				-.07 (.09)	-.02 (.11)
Home Office Lighting (Baseline: Average)					
Natural				.02 (.09)	.01 (.11)
No Natural				-.01 (.17)	-.11 (.21)
Home Office Ventilation (Baseline: None)					
Mechanic				-.07 (.16)	-.17 (.18)
Manual				-.08 (.14)	-.11 (.16)
Home Office surface (m^2)				-.03 (.03)	-.05 (.03)
Real-estate value (x€1,000)					-.01 (.04)
Address-density (per kilometer radius)					-.06 (.07)
Urbanicity (Baseline: Extremely high)					
High					-.10 (.13)
Moderate					-.06 (.16)
Low					.0003 (.18)
None-Urban					-.04 (.21)
Observations	1,002	1,002	1,002	956	734
R2	.25	.27	.30	.30	.30
Adjusted R2	.23	.25	.28	.27	.26
Residual Std. Error	.88 (df = 972)	.87 (df = 971)	.85 (df = 968)	.85 (df = 915)	.83 (df = 687)
F Statistic	11.41*** (df = 29; 972)	12.16*** (df = 30; 971)	12.79*** (df = 33; 968)	10.01*** (df = 40; 915)	6.50*** (df = 46; 687)

Note. *p<0.1, **p<0.05, ***p<0.01.

Table 4.9 Complete Regression Table of Burnout Propensity

	Dependent variable: Burnout Propensity				
	(1)	(2)	(3)	(4)	(5)
Desk & Chair	-.14 (.04)***		-.10 (.04)**	-.11 (.05)**	-.16 (.06)***
Screen & Hardware	-.09 (.05)*		-.04 (.05)	-.03 (.05)	-.001 (.06)
WiFi	-.12 (.04)***		-.07 (.04)*	-.07 (.04)*	-.05 (.05)
Temperature		-.07 (.04)*	-.04 (.04)	-.06 (.04)	-.06 (.06)
Air Quality		-.11 (.04)**	-.08 (.04)*	-.09 (.04)**	-.07 (.05)
Lighting		-.05 (.04)	-.02 (.04)	.01 (.04)	.02 (.05)
Noise		-.13 (.04)***	-.10 (.04)***	-.09 (.04)**	-.09 (.05)**
Age (years)	-.01 (.003)***	-.01 (.003)***	-.01 (.003)***	-.01 (.003)**	-.01 (.003)*
Income (Baseline: Modal)					
Mininum wage (less than 11,000)	.63 (.26)***	.56 (.27)***	.56 (.27)***	.57 (.28)***	.46 (.32)*
below modal (11-23k)	.26 (.14)**	.22 (.14)*	.23 (.14)**	.27 (.14)**	.24 (.17)*
1-2x modal (34-56k)	-.01 (.08)	-.02 (.08)	-.02 (.08)	-.02 (.08)	-.05 (.10)
2x modal or more (56k)	.13 (.10)	.12 (.10)	.13 (.09)	.12 (.10)	.07 (.11)
don't know/ don't want to say	.21 (.11)**	.20 (.11)*	.21 (.11)**	.23 (.11)**	.24 (.13)**
Female	.01 (.07)	.09 (.07)	.05 (.07)	.09 (.07)	.03 (.08)
Household Members	-.10 (.04)**	-.11 (.05)**	-.10 (.04)**	-.09 (.05)*	-.08 (.05)
Children Home during Office Hours (baseline: no children)					
Always	.42 (.22)**	.42 (.22)**	.39 (.22)**	.35 (.21)*	.42 (.24)*
Sometimes	.06 (.10)	.07 (.10)	.06 (.10)	.02 (.10)	-.02 (.11)
Never	-.24 (.09)**	-.23 (.09)**	-.23 (.09)**	-.25 (.09)**	-.26 (.11)**
Partner Home during Office Hours (baseline: no Partner)					
Always	.08 (.09)	.08 (.09)	.08 (.09)	.09 (.10)	.13 (.11)
Sometimes	-.07 (.10)	-.05 (.09)	-.07 (.09)	-.03 (.10)	-.11 (.11)
Never	-.14 (.09)	-.09 (.09)	-.11 (.09)	-.10 (.10)	-.10 (.11)
Pet (Baseline: No pets)					
Dog	.15 (.08)**	.16 (.08)**	.17 (.08)**	.19 (.08)**	.17 (.10)*
Cat	-.002 (.07)	-.01 (.07)	.001 (.07)	.01 (.07)	-.02 (.08)
Company size (Baseline: 0-5)					
5-15	-.03 (.15)	.02 (.14)	.02 (.14)	.02 (.15)	.04 (.19)
15-50	-.15 (.14)	-.10 (.13)	-.10 (.13)	-.06 (.14)	-.10 (.18)
50+	.16 (.14)	.18 (.13)	.18 (.13)	.20 (.14)	.16 (.17)
Work Sector (Baseline: Governmental)					
Yes, non-governmental	.11 (.07)	.07 (.07)	.08 (.07)	.09 (.08)	.15 (.08)*
Yes, temp/ on-call worker	-.05 (.20)	-.15 (.19)	-.11 (.19)	-.12 (.22)	-.10 (.25)
Yes, self-employed	-.09 (.13)	-.14 (.13)	-.10 (.13)	-.05 (.14)	-.09 (.16)
Contract hours (Baseline: Full time (36+)					
20-35 hours	.01 (.07)	.02 (.07)	.02 (.07)	-.02 (.07)	.03 (.08)
12-19 hours	-.15 (.14)	-.14 (.13)	-.14 (.14)	-.16 (.14)	-.03 (.15)
less than 12 hours	-.33 (.14)**	-.25 (.15)	-.29 (.15)*	-.32 (.16)*	-.49 (.17)**
Work suitable to perform from home	-.01 (.03)	-.03 (.03)	-.01 (.03)	-.01 (.03)	.01 (.04)
Home Office Floor plan (Baseline: Average)					
Open				.07 (.10)	.04 (.11)
Closed				.07 (.09)	.04 (.11)
Home Office Lighting (Baseline: Average)					
Natural				-.16 (.10)	-.15 (.12)
No Natural				.10 (.19)	-.01 (.20)
Home Office Ventilation (Baseline: None)					
Mechanic				.17 (.19)	.14 (.22)
Manual				.03 (.17)	.01 (.19)
Home Office surface (m²)				.03 (.03)	.04 (.04)
Real-estate value (x€1,000)					.01 (.04)
Address-density (per kilometer radius)					-.03 (.06)
Urbanicity (Baseline: Extremely high)					
High					-.02 (.13)
Moderate					-.02 (.16)
Low					-.15 (.16)
None-Urban					-.14 (.20)
Observations	1,002	1,002	1,002	956	734
R2	.18	.19	.21	.21	.21
Adjusted R2	.16	.17	.18	.17	.16
Residual Std. Error	.92 (df = 972)	.91 (df = 971)	.91 (df = 968)	.90 (df = 915)	.91 (df = 687)
F Statistic	7.60*** (df = 29; 972)	7.71*** (df = 30; 971)	7.70*** (df = 33; 968)	6.06*** (df = 40; 915)	3.93*** (df = 46; 687)

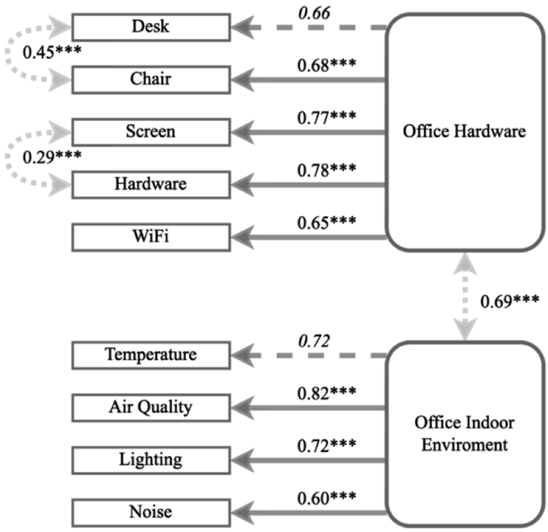
Note. **p*<0.1, ***p*<0.05, ****p*<0.01.

Table 4.10 Complete Regression Table of Willingness to continue working from home

	Dependent variable: Willingness to continue working from home				
	(1)	(2)	(3)	(4)	(5)
Desk & Chair	.14 (.04)***		.11 (.04)**	.11 (.05)**	.09 (.06)*
Screen & Hardware	.0004 (.05)		-.03 (.05)	-.02 (.05)	.05 (.06)
WiFi	.12 (.04)***		.07 (.04)**	.08 (.04)**	.06 (.04)
Temperature		.12 (.04)***	.09 (.04)**	.09 (.04)**	.07 (.05)
Air Quality		.02 (.05)	.004 (.05)	-.01 (.05)	.01 (.06)
Lighting		.08 (.04)**	.06 (.04)*	.08 (.04)*	.03 (.05)
Noise		.05 (.04)	.03 (.04)	.01 (.04)	-.002 (.05)
Age (years)	.01 (.003)***	.01 (.003)***	.01 (.003)***	.01 (.003)***	.01 (.003)**
Income (Baseline: Modal)					
Mininum wage (less than 11,000)	-.17 (.19)	-.14 (.20)	-.13 (.19)	-.06 (.20)	-.01 (.24)
below modal (11-23k)	.03 (.11)	.08 (.11)	.07 (.11)	.08 (.11)	.11 (.12)
1-2x modal (34-56k)	-.11 (.08)	-.10 (.08)	-.11 (.08)	-.10 (.08)	-.09 (.10)
2x modal or more (56k)	-.16 (.09)*	-.14 (.09)	-.14 (.09)	-.15 (.10)	-.08 (.11)
don't know/ don't want to say	-.02 (.10)	-.005 (.10)	-.01 (.10)	-.02 (.10)	.01 (.12)
Female	.09 (.06)	.05 (.06)	.07 (.06)	.06 (.07)	.05 (.08)
Household Members	-.004 (.05)	.01 (.04)	.004 (.04)	.01 (.05)	-.01 (.05)
Children Home during Office Hours (baseline: no children)					
Always	-.23 (.19)	-.24 (.18)	-.21 (.19)	-.28 (.20)	-.32 (.23)
Sometimes	-.10 (.10)	-.12 (.09)	-.10 (.10)	-.13 (.10)	-.07 (.11)
Never	.03 (.09)	.03 (.09)	.03 (.09)	.03 (.09)	-.02 (.11)
Partner Home during Office Hours (baseline: no Partner)					
Always	.07 (.09)	.09 (.09)	.09 (.09)	.11 (.10)	.07 (.11)
Sometimes	.12 (.09)	.11 (.09)	.11 (.09)	.14 (.09)	.11 (.11)
Never	.14 (.09)	.11 (.09)	.12 (.09)	.15 (.09)	.08 (.11)
Pet (Baseline: No pets)					
Dog	.09 (.07)	.09 (.07)	.08 (.07)	.07 (.07)	.03 (.09)
Cat	-.04 (.07)	-.04 (.06)	-.04 (.06)	-.04 (.07)	-.02 (.08)
Company size (Baseline: 0-5)					
5-15	-.25 (.15)	-.28 (.15)*	-.27 (.15)*	-.25 (.16)	-.15 (.18)
15-50	-.31 (.14)**	-.34 (.13)**	-.34 (.14)**	-.29 (.14)**	-.23 (.17)
50+	-.36 (.12)***	-.37 (.12)***	-.37 (.13)***	-.34 (.13)**	-.25 (.16)
Work Sector (Baseline: Governmental)					
Yes, non-governmental	.05 (.07)	.08 (.07)	.08 (.07)	.08 (.07)	.07 (.09)
Yes, temp/ on-call worker	-.03 (.14)	.03 (.15)	-.01 (.14)	-.09 (.16)	-.12 (.21)
Yes, self-employed	.08 (.13)	.09 (.14)	.07 (.14)	.10 (.14)	.04 (.16)
Contract hours (Baseline: Full time (36+)					
20-35 hours	.01 (.07)	.01 (.07)	.01 (.07)	.03 (.07)	.05 (.08)
12-19 hours	-.05 (.14)	-.09 (.14)	-.08 (.14)	-.04 (.15)	.02 (.16)
less than 12 hours	-.01 (.14)	-.09 (.14)	-.06 (.14)	-.05 (.15)	-.24 (.18)
Work suitable to perform from home	.32 (.03)***	.34 (.03)***	.33 (.03)***	.32 (.03)***	.34 (.04)***
Home Office Floor plan (Baseline: Average)					
Open				-.02 (.09)	-.03 (.10)
Closed				-.12 (.09)	-.16 (.10)
Home Office Lighting (Baseline: Average)					
Natural				.01 (.09)	.08 (.11)
No Natural				.21 (.18)	.21 (.21)
Home Office Ventilation (Baseline: None)					
Mechanic				.22 (.18)	.18 (.20)
Manual				.04 (.16)	.03 (.18)
Home Office surface (m^2)				.01 (.03)	-.02 (.04)
Real-estate value (x€1,000)					-.01 (.03)
Address-density (per kilometer radius)					-.02 (.06)
Urbanicity (Baseline: Extremely high)					
High					.21 (.12)*
Moderate					.01 (.15)
Low					.02 (.17)
None-Urban					.05 (.19)
Observations	1,002	1,002	1,002	956	734
R2	.24	.25	.26	.26	.26
Adjusted R2	.22	.22	.23	.23	.21
Residual Std. Error	.88 (df = 972)	.88 (df = 971)	.88 (df = 968)	.88 (df = 915)	.88 (df = 687)
F Statistic	10.65*** (df = 29; 972)	10.63*** (df = 30; 971)	10.13*** (df = 33; 968)	8.03*** (df = 40; 915)	5.24*** (df = 46; 687)

Note. *p<0.1, **p<0.05, ***p<0.01.

Figure 4.5 Structural Equation Model Latent Variables Loading and Covariance



Note. *p<0.05, **p<0.01, ***p<0.001

Chapter 5

Turning up the Heat

The impact of indoor temperature on selected cognitive processes and the validity of self-report.*

5.1 Introduction

Performance at work is influenced by many factors, such as individual characteristics, leadership, work pressure, incentive schemes, and corporate structure (Hermalin & Weisbach, 1991; Perry & Porter, 1982; Wageman & Baker, 1997). The physical climate of the workplace is often overlooked as an important factor influencing performance. And when it is mentioned, the dominant strain of research focuses on comfort, through self-report on physical aspects of the environment and their effect on human performance. This is remarkable, as office buildings have been undergoing rigorous innovations throughout recent decades (for instance, Vermeulen & Hovens, 2006). Developments in the quality of insulation, ventilation, and air-conditioning are effectively changing the indoor environment to which workers are exposed. These innovations are typically motivated by effects on building efficiency and/or worker comfort, but while there is ample research highlighting the effects of increased energy efficiency on building resource consumption (Eichholtz et al., 2019; Pérez-Lombard et al., 2008), the link between changes in indoor environmental conditions and human performance remains a topic of debate (MacNaughton et al., 2017; Satish et al., 2012; X. Zhang et al., 2017).

Research regarding the impact of indoor environment on worker performance is hampered by the fact that high-skilled performance measures at work are difficult to obtain directly, and are hard to compare between disciplines. For example, Zivin and Neidell (2012) show that pear-pickers' performance suffers from exposure to bad environmental quality conditions. However, the output of highly skilled workers who face cognitively demanding tasks

– such as academics, managers, doctors, or investors – lacks such direct outcome measure. It is exactly this type of high-skilled workers who spends considerable time in confined offices or meeting rooms, subject to specific indoor climate conditions. Parsons (2014) notes that individual factors often dominate performance outcomes, making it even more challenging to compare productivity between workers. Moreover, any output that is measurable is not easily traced back to a quantifiable time period of exposure to the physical indoor climate.

To circumvent the challenge to correctly assess human performance, research has shifted from measuring performance to comfort (Bluyssen, 2013). The implicit expectation is that when the climate is rated as "comfortable", productivity increases. Comfort measures are an attractive proxy for productivity and performance, as they are easily and inexpensively assessed by self-report. Comfort could be treated as a measure of interest on its own (for instance, Nakamura et al., 2008), but whether self-assessed comfort levels are indeed an accurate proxy for performance remains an open question. Psychological research repeatedly suggests self-reported introspection into one's own subjective experience and emotions to be unreliable (Engelbert & Carruthers, 2010).

In this paper, we assess the effect of indoor environmental conditions on human performance, by investigating decision processes. Tversky and Kahneman (1974), amongst others, distinguish decision making as "intuitive" and "rational" processes. Automated, intuitive rules of thumb, or heuristics, are "quick and dirty" and applied without much effort. The rational processes need more time and cognitive resources, are only scarcely applied, and are also associated with high decisional quality. A mainstream application of the interplay between these fast and rational or effortful processes is the default-interventionist approach (Evans, 2007). It stipulates that the effortful processes can intervene in the fast heuristics, when a wrongful application (a bias) in a given context is detected. Thus, whenever the effortful processes are hampered, for instance due to cognitive constraint resulting from environmental factors, increased bias-susceptibility generally lowers overall decisional quality (Gawronski & Bodenhausen, 2006; Muraven & Baumeister, 2000). In other words, we expect that bias detection and correction will (partially) suffer due to cognitive constraint in effortful processes capacity following temperature stress.

5.2 Literature

5.2.1 Temperature and Cognition

Psychological and neurological research has attempted to identify the effects of temperature on cognitive functions. We elaborate on two relevant findings.

The most profound and general finding is that cognitive capacity is lowered by adverse temperature conditions. Wright, Hull and Czeisler (2002) find that changes in the temperature of the body and brain are correlated with changes in performance, such that deviating temperatures from the internal optimal will worsen performance. Shibasaki, Namba, Oshiro, Kakigi and Nakata (2017) show that neurological inhibition processes suffer from heat stress. In decision-making, executive and inhibition processes coordinate which stimuli to act on (execute) and which not (inhibit). Both these biological processes are found to be weaker under heat stress. Van Ooijen, Van Marken Lichtenbelt, Van Steenhoven and Westerterp (2004) suggest that temperature could influence mental performance as a result of fatigue. This view is similar to the theoretical concept of mental depletion, the cognitive model stipulating limited mental "control" resources for self-regulation (R. Baumeister et al., 1998). Mental depletion often results in more instinctive behavior (such as aggression; Van Lange et al., 2017). In general, when external stimuli overstimulate, concentration and performance become more costly (MacLeod, 1991).[58] Indeed, Cheema and Patrick (2012) show that temperature generally lowers cognitive performance, but not for people who were already mentally depleted at the start of the task. Although mental depletion is debated (Carter et al., 2015; Hagger et al., 2016), the general notion of negative cognitive performance effects after enduring strain on mental capacity seems to be a common denominator in ongoing self-regulation discussions (R. Baumeister et al., 2007; Cunningham & Baumeister, 2016; Hockey, 2013; Inzlicht et al., 2021; Lin et al., 2020).

The second key finding of research on temperature and cognition is that not all mental processes are affected equally. Lowered cognitive capacity appears theoretically very close to behavioral fatigue. However, it is important to understand that these two concepts are fundamentally and hierarchically distinct. When discussing behavioral fatigue, we consider a general lowering of behavioral activity (i.e., a "global" effect). Decrease of cognitive capacity does not have a general uniform effect, but is depending on the neurological area that suffers

[58] The distraction due to discomfort and the active act to ignore this distraction can drain additional resources from available mental capacity. However, the majority of the research previously described sees a loss of performance independent of awareness, suggesting that awareness of discomfort alone does not fully explain the decrease in performance. Additionally, the temperature dissatisfaction levels in our experiment do not reach extreme levels, suggesting against high levels of rumination during the task. Therefore, we argue that the physiological capacity limitations from compensating the effect heat has on the body and its processes is more profound and is our main focus.

most (i.e., a "local" effect). Lan, Lian, Pan and Ye (2009) found performance to decrease with adverse temperatures, but the effects differ across tasks.

In sum, it is clear that temperature has a general, or global, effect on cognition and cognitive performance, and that some local effects can be identified as well.

5.2.2 Temperature and Intuition

The literature review by Hancock and Vasmatzidis (2003) suggests that high capacity and complex mental processes are more profoundly affected by temperature than automated processes. Automated tasks rely on a strong and fast relation between stimulus and response, making them less susceptible to mental constraints (Kahneman, 1973). Automated tasks are part of system I in Kahneman's cognitive framework – also known as the intuitive system. They rely on intuition and on simple rules of thumb that are learned and are often successfully applied to predictable situations. System II is slow and costly on mental resources, but is generally associated with high-quality decision making.

Cognitive capacity and cognitive control are highly correlated (Engle & Kane, 2003), and the latter has also been found to be affected by temperature. Shibasaki, Namba, Oshiro, Kakigi and Nakata (2017) show that neurological inhibition processes suffer from heat stress. In decision making, inhibition and executive processes coordinate to achieve an optimal solution. As such, the effect of heat on performance can be twofold: not only do higher-order complex tasks suffer more than simple automated tasks (Grether, 1973), but wrongful application of an automated process or application of a wrong automated process might also be less likely to be corrected. In other words, even when the direct effect of heat on simple and automated processes is not evident (as stated by F. Zhang & De Dear, 2017), the outcome can still suffer in quality due to the lack of high order process intervention. Indeed, Hancock and Vasmatzidis (1998) found that highly skilled operators suffer less from performance decrease under heat stress, and they argued that this is most likely a result of performance depending on automated internalized processes.

The cognitive framework of Tversky and Kahneman leads to relevant predictions when we apply the findings of temperature on task complexity and intuition. The interaction found between temperature and automated tasks and task complexity suggests that system I could be less affected than system II. The default-interventionist approach (Evans, 2007) stated that both systems work parallel to each other, and system II generally attempts to identify mistakes made by system I and intervenes if necessary. Recent advances in this field suggest that logical conclusions also manifest intuitively (De Neys & Pennycook, 2019). In this view, deliberation by system II is activated only when both the heuristic and logic intuition are of similar strength and conflicting. Thus, a correct response on the CRT, for instance, does not need deliberation when the logic intuition is stronger than the heuristic intuitive. For both views, however, the wrongful application of heuristics would be more prevalent when the controlling function of system II would fail as a consequence of the heat stress.[59]

We therefore expect that the distinct effect that heat has on cognition can be (partially) captured by the Kahneman framework. Recent research has investigated the effect on cognitive reflection (Chang & Kajackaite, 2019), but to date, no study has extended this investigation to the specific behavioral biased outcomes stemming from a predisposition to overly adhere to intuitive decision strategies. Although the CRT is highly correlated with specific behavioral biases, we test the effect of heat on bias sensitivity for an array of specific well-known biases directly. To our knowledge, no attempts have been made to distinguish the effects of heat on behavior and cognition using this approach.

5.2.3 Temperature and Risk

Evidence suggests that temperature has a direct effect on the willingness to take risk. Wang (2017) shows that people making trading decisions will pursue high-risk high-yield options compared to a control condition. Some indirect evidence on aggression also suggests that risky behavior could follow from loss of control through the same channel. For instance, solely increasing the temperature makes people subjectively rate other people in the room to be more hostile (Anderson et al., 2000). Cao & Wei (2005) hypothesize that aggression leads to increased risk behavior. Denson, DeWall and Finkel (2012) conclude that it is the loss of self-control that increases aggression. Finally, Frey, Pedroni, Mata, Rieskamp and Hertwig (2017) show self-control to be predictive of various risk behavior outcomes. Overall, we expect the same channel that increases system I dependency will also increase risk-taking behavior.

5.2.4 Temperature and Gender

Many individual characteristics mediate the effect heat has cognition, however, the heterogeneous gender-related differences stand out.[60] Biological research (Kingma & Van

[59] We discuss the implications of this renewed model in light of our results in the limitations section.
[60] We extensively discuss the potential influence of other individual characteristics in the limitation section.

Marken Lichtenbelt, 2015), metabolic research (Byrne et al., 2005), and psychological empirical research (Wyon, 1974) show that hot temperatures have a distinctly different effect on women as compared to men. The most profound example of this distinction and its neglect in the past decade is the temperature comfort level. The 'default' room temperature level of 21° C seems mainly based on male preferences (Kingma & Van Marken Lichtenbelt, 2015) Indeed, anecdotal evidence suggests that women perform better at slightly higher default room temperatures (Chang & Kajackaite, 2019).

As such, finding the effects of adverse temperature on cognition would be incomplete without taking gender-specific preferences into consideration. Without correcting for gender, female preference or tolerance for higher temperatures might influence the overall findings regarding the effect of adverse temperatures on performance. Given that women show a preference for somewhat higher temperatures, women will rate identical absolute temperature increases (subjectively) as less adverse as compared to men. Performance for women might thus also be expected to be less affected by heat.

5.2.5 This study

We hypothesize that heat exposure will decrease cognitive performance such that biased behavior will be more prominent, as rational correction will require more effort under heat stress. Heat is a salient factor in the working environment and workers can often elicit control over temperature themselves, making the relevance of our results apparent and immediately applicable. Moreover, by testing detectable temperature differences in each condition, we are able to assess the accuracy and thus relevance of self-reported comfort measures for in future research.

Additionally, we investigate the effect of heat on risk behavior. Through the same channel, we expect that a combination of lack of effortful control and bodily discomfort will increase risk behavior. This would be in line with aggression studies (for instance, American football players commit more aggressive fouls; Craig et al., 2016). We test both the general self-reported risk attitude, which has generally been claimed to be a rather stable character trait, unaffected by heat (Dohmen et al., 2011a), and actual risk behavior, which we expect to increase following indoor temperature manipulation (see, for example, Wang, 2017).

Our experimental design has several key advantages over current practices in the literature. First, we actively strive to control a variety of factors influencing the physical experience of the environment. That is, we pre-expose all participants to the temperature manipulation for a defined adjustment period of one hour before starting the tasks. All participants are wearing similar clothing provided specifically for the experiment. We further control for the outdoor temperature of the period before testing. Second, we keep all other indoor climate factors constant. For instance, we manipulate the temperature while keeping air ventilation levels unchanged. As a result, CO_2 levels, noise, lighting, and air refreshment are equal between manipulations. Some recent experiments manipulated temperature by opening and closing windows, without controlling for CO_2 and fine particles between groups, and are therefore unable to isolate the effect of just temperature on task performance (X. Wang, 2017).

5.3 Method

5.3.1 Experimental conditions and design

We designed a controlled experiment to measure the effect of heat on decision quality. We employed a stratified random sampling method to recruit a total of 257 participants with an average age of 21.57 (SD = 2.41) years old using the Maastricht University Behavioral Experimental Economics laboratory database. Stratification ensures an equal gender distribution amongst manipulation groups. The final sample allows for a 10% deviation of gender within groups. All participants were proficient in reading and writing of the English language. Participants are randomly distributed to either the control or the experimental condition.[61] This between-subject design used temperature as the main independent variable. Given the clear gender differences in the temperature effect on performance and satisfaction in the literature, gender is the secondary independent variable in our analysis.

Participants were exposed to a controlled physical environment with either a hot temperature (28° C) or a neutral temperature (22° C). The decision for 28° C is derived from the body of literature focused on temperatures below 29° C / 85° F (for an overview, see Hancock & Vasmatzidis, 2003). More specifically, previous research repeatedly showed an effect of hot temperature on performance on neurobehavioral test at 27–28° C (Lan et al., 2009; Lan & Lian, 2009).[62] In these conditions, a battery of validated tests included cognitive reflection tasks, a heuristics battery, lottery risk tasks, and self-reported risk preferences. Additionally,

[4] Appendix Table 5.4 Panel B summarizes individual characteristics per condition.

[62] As we discuss in the limitations section, we acknowledge that higher temperatures could show more profound effects. However, the goal of this paper is to generalize our results to the professional workforce. We argue that the relevance of excessive temperatures upwards of our threshold of 27–28° C will be exponentially decreasing with each increase in degree Celcius. Temperature measurements in real-life settings show repeatedly naturally occurring temperature variations of 28° C within one standard deviation of the mean, but rarely above 29° C (Künn et al., 2019; Zivin et al., 2018).

participants state their personal comfort levels and their subjective estimation as to what extent the environment influences their performance on the battery of tasks. The experiment was programmed using Qualtrics Software (Qualtrics, Provo, UT) and executed at the Behavioral Experimental Economics lab facilities at Maastricht University in the Netherlands. The laboratory is approximately 5 meters wide and 20 meters long. In this room, there are 33 cubicles (approx. 1.0 meter by 1.5 meters), all including a computer and table, which are closed off by shutters. All participants are tested in groups varying between 25 and 30 participants per group. Air quality is controlled using a climate system that holds the air refreshment rate constant.[63] The control condition of 22 ° C is reached running only the climate system. The "hot" condition of 28 ° C is reached using five 3kW industrial heaters, each with a 115m^3 capacity. During the experiment, four heaters maintain a constant temperature. Manual adjustments to the thermostats of the individual heaters ensure a stable temperature. All heaters also ran without heating during the control condition, such that the noise produced by the heaters is constant between conditions.[64]

All participants were subject to strict clothing prescriptions. These requirements ensure that all participants have a similar physical experience of the heat. For instance, the possibility to remove layers of clothing could increase heterogeneity in the experienced heat within and between conditions. All participants are asked to wear long jeans. To fully ensure homogeneity, we provide all participants with long-sleeved black polyester thermoshirts. Participants are not allowed to wear anything underneath these shirts.[65]

Participants arrived in the laboratory at 11 AM, one hour before the start of the actual experiment. This adaption time ensured that all participants experience the indoor climate similarly, independent of the outdoor temperature or previous activity. During this adaption time, the temperature was kept at the same levels as during the experiment. After one hour, the test battery automatically started. All tasks were completed in English. Each task was presented to each participant only once. We did not impose a time schedule for the different tasks. The average completion time was roughly 45 minutes. Moreover, the outdoor temperature was

[63] See Appendix Table 5.4 Panel A for an overview of the average CO2 and humidity per condition.

[64] Although individual preferences and satisfaction regarding illumination and acoustics will differ, the fact that all participants were exposed to the same conditions leads us to conclude that there is no objective reason why we would find a significant difference between the reported satisfaction on either of these variables between the control and manipulation group, on average.

[65] Women are allowed to wear bras underneath. We estimate that the clothing insulation value of all the subjects' ensemble is around 0.65 clo, on average (based on American Society of Heating, 2017). However, we note that our main purpose is to minimize variation between subjects and conditions. Therefore, the relative clo value of the clothing between groups is more relevant for the interpretation of the results than the absolute value of the ensemble.

measured on all testing days and compared between conditions (Appendix Table 5.4 Panel A provides an overview of the indoor temperature during task and adaption, as well as the outdoor temperature between conditions.) The tasks were given in the order in which they are presented in Section 5.3.2. All tasks were presented to each participant only once.

5.3.2 Dependent measures

5.3.2.1 Performance measures

Cognitive Reflection Task. The classic Cognitive Reflection Task (CRT) by Frederick (2005) measures participants' propensity to rely on intuition or rational thinking. The test consists of three questions, of which each question has a salient intuitive answer and a correct rational answer. Each of these questions are scored with 1 for a correct response or 0 for an incorrect response. The score for this task is the number of correctly answered questions, such that the score of the CRT lies between 0 (no correct answers) and 3 (all answers correct). Although this test is often used, Bialek & Pennycook (2017) find that multiple exposure does not reduce its validity.

Cognitive Reflection Task Expansion. To increase the probability of capturing the distinction between intuitive and rational thinking in our sample, we added an expansion of the original CRT. This test (from Toplak et al., 2014) consists of three additional items, following the same structure. It is highly correlated to the original CRT.

Heuristics Battery. The heuristic bias task battery by Toplak, West and Stanovich (2011) includes various questions about well-known economical biases. We select ten questions from this battery concerning casual base rate neglect, sample size problems, sensitivity towards regression to the mean, framing bias, outcome bias, the conjunction fallacy, probability matching, ratio bias, methodological reasoning, and the covariation problem.[66] Each of these questions are scored with 1 for a correct response or 0 for a biased and thus wrong response. The resulting score on this battery is thus between 0 and 10 points (M = 6.32, SD = 2.16), in line with the original authors.

[66] For an overview of these tasks, see Toplak et al. (2011)

5.3.2.2 Risk measures

Risk Elicitation Task. The first measure of risk assessment is aimed at inducing or eliciting actual risk behavior at the time of the experiment. Similar to the original task of Holt and Laury (2002) we showed the participants nine choices between two sets of lotteries. The first lottery is of relatively low risk, where both the high and low payout options diverge only minimally (€6 versus €4.80, respectively). The second lottery can be considered high risk, as there is a strong divergence between the high (€11.55) and low (€0.30) payout option. For each consecutive choice, the probability of the high payout in both lotteries increases with 10%, such that in the first choice the probability of the high payout for each lottery is 10% and in the ninth and final choice this probability has become 90%. Note that the expected payout of the high-risk lottery surpasses the payout of the low-risk lottery from step 5 onwards (since then the expected payout is €5.93 for the high-risk versus €5.40 for the low-risk lottery). Participants are scored on a scale from 1–10, where the score reflects the switching point of the participants. Score 1 indicates a sustained preference for the high-risk lottery, labelling them as "risk-loving". A score of 5 implies risk-neutral behavior, as participants follow the switching point in which both measures are equivalent. A score of 10 is assigned when participants never switch to the high-risk lottery. We label these participants as "risk averse". Depending on the risk preference, all scores are considered rational, as even in step 1 or 9 there is still a 10% probability of a high win or loss, respectively. This lottery is incentivised, and participants are told that one of the lottery choices will be played at the end of the questionnaire. The outcome of their chosen lottery will be added to their total reimbursement. To make this incentive at least 25% of the total reimbursement, the lottery outcomes are multiplied by a factor from the original (Holt & Laury, 2002). Participants who switched their choice of lottery more than once were excluded from the sample; 34 observations were thus excluded (16 male, 18 female).[67]

Risk Attitude Task. In addition to a risk elicitation task, we asked participants how risk-loving they perceive themselves to be, both in general and on specific domains. Participants rated themselves on a 10-point scale, with the lowest score being risk-averse, and the highest score labelled fully prepared to take risk. First, all participants state to what extent they are willing to take risk or avoid taking risk generally as a person. Second, their willingness to take or avoid risk are specified for the following domains: driving, financial matters, leisure and sport, their

[67] These participants were only excluded for the risk elicitation analysis. We found no indication that these participants were structural outliers throughout the task battery thus we did not conclude that their inconsistency in the risk elicitation task invalidated their scores for all other tasks.

occupation, health, and faith in other people. This approach has been extensively validated and found to correlate with actual risk behavior (Dohmen et al., 2011; Falk et al., 2016).

5.3.2.3 Indoor climate satisfaction

Self-reported Indoor Climate Satisfaction and Hindrance. Self-reported indoor environmental satisfaction was assessed by adapting the occupant indoor environment quality survey developed by Berkeley's Centre for the Built Environment (Huizenga et al., 2006). For temperature, air quality, noise, and lighting, all participants are asked to rate their satisfaction level on a scale from 1 to 7. Additionally, for all these factors, participants are asked to what extent they perceive it as hindering or supporting their ability to answer the questions in the questionnaire on a similar 7-point scale. Note that at this time, participants completely the performance tasks. To what extent the climate influenced their performance was therefore not predictive, but reflective. The scores are recoded such that a score of 7 indicates that the factor fully supports their ability, and a score of 1 indicates that the factor fully hinders their ability to answer the questionnaire. We label the totality of these factor-specific measures "satisfaction measures". In the analysis, we control for multiple testing.[68]

5.3.2.4 Additional checks

CRT multiple exposure check. After the three performance tasks (e.g., original CRT, extended CRT, and the Heuristics battery), all participants were asked to indicate whether they recognize any if these questions and if yes, whether they also remember the correct answer. These questions are scored by 1 – yes, 2 – no, or 3 – unsure.

Clothing check. All participants were asked to indicate whether they are indeed wearing the thermoshirts provided by the experimenter.[69] On a Likert-scale of 1 (bad) to 7 (good), participants indicate the fit, length, and the comfort of the shirt. Additionally, we ask to what extent the shirt influences the performance on the tasks using the same scale.

Temperature. To be able to check for climate adjustment effects, three questions assessed the current and past climate experienced by the participants as well as their climate preference.

[68] Multiple testing correction is applied for all 10 conditions using the Benjamini & Hochberg procedure (Benjamini & Hochberg, 1995), see Appendix Table 5.6. This procedure aims to control the false discovery rate whilst preserving relatively higher power compared to more conservative procedures (e.g., Bonferroni correction; (Thissen et al., 2002).

[69] One of the participants indicated to be allergic to the fabric of the thermoshirts, and was thus asked to wear a similar (long-sleeved) shirt. All other participants wore the thermoshirts provided by the experimenter.

Specifically, participants were asked to state in which country they grew up (most time spend until your 18th birthday), in which country they lived for the majority of the last five years, and what their preferred thermostat setting is (in degrees Celsius) in winter.

5.3.3 Incentives payoff

The payout was determined by adding the outcome of the preferred lottery of the risk elicitation task to the standard endowment of €15. The participants were told that for one of the steps, their chosen lottery will be played, but do not know which step this will be. The Qualtrics Internal Randomizer was used to draw an outcome (50/50 allocation) for the lottery chosen by the participant at step 5. The outcome was displayed at the end of the questionnaire. For the whole sample the average expected payoff of the risk task is 27% of the total payoff (with mean €5.98). No other performance tasks were incentivised, as these specific tasks are found not to be affected by incentives (Brañas-Garza et al., 2019).

5.3.4 Statistical approach

To investigate statistical significance of the variables of interest, we ran mean comparison tests between the two manipulation conditions. Specifically, we conducted independent samples t-tests using STATA software (StataCorp., 2017). In situations when normality violations are detected (using Shapiro-Wilk normality tests), we tested for significance using Mann-Whitney U (Wilcoxon rank-sum) tests. For all results, we state whether parametric or nonparametric procedures are reported. Additionally, we apply the Benjamini & Hochberg procedure (Benjamini & Hochberg, 1995) as multiple testing correction when required.

5.4 Results

5.4.1 Descriptives and Condition Manipulations

The recorded sample consists of 257 students ranging from 17 to 31 years old, of which 53.5% are female (see Appendix Table 5.3).[70] The recorded indoor and outdoor climate conditions are reported in Appendix Table 5.4. The average temperature in the control condition was 22.4° C and in the hot condition 28.3° C. Levels of indoor CO_2, outdoor temperature of

[70] The sample shows a average self-reported math proficiency of 63 on a scale from 0 to 100

each test day during the morning, and outdoor temperature of the past three days do not differ significantly between manipulations.

5.4.2 Satisfaction measures

We first present the climate satisfaction measures in Table 5.1. Looking at the first column, it is confirmed that temperature (*d*= 0.77) and air quality (*d*= 1.53) are significantly less satisfactory in the hot condition. Additionally, both are predicted to hinder the performance on the performance measures. This confirms the notion that the high-temperature manipulation is considered uncomfortable.

Table 5.1 Main Results of Indoor Variables

				Men			Women		
	Control	Hot	*p*-value	Control	Hot	*p*-value	Control	Hot	*p*-value
Panel A: Satisfaction									
Temperature	4.66 (1.57)	3.50 (1.45)	.00***	5.13 (1.53)	3.05 (1.29)	.00***	4.25 (1.49)	3.90 (1.49)	.16
Air Quality	5.35 (1.18)	3.54 (1.41)	.00***	5.32 (1.23)	3.38 (1.39)	.00***	5.38 (1.15)	3.67 (1.44)	.00*
Light	5.33 (1.46)	4.95 (1.64)	.07	5.50 (1.55)	5.57 (1.03)	.56	5.19 (1.39)	4.42 (1.88)	.02*
Noise	5.36 (1.43)	5.57 (1.42)	.18	5.42 (1.51)	5.58 (1.39)	.58	5.30 (1.36)	5.55 (1.46)	.18
Clothing	5.71 (1.36)	5.55 (1.27)	.14	5.62 (1.37)	5.08 (1.33)	.02*	5.80 (1.37)	5.96 (5.96)	.81
Panel B: Hinderance									
Temperature	4.68 (1.54)	3.40 (1.49)	.00***	5.27 (1.25)	3.05 (1.25)	.00***	4.17 (1.60)	3.71 (1.62)	.07
Air Quality	5.07 (1.23)	3.71 (1.45)	.00***	5.03 (1.25)	3.65 (1.23)	.00***	5.10 (1.22)	3.75 (1.63)	.00*
Light	5.02 (1.55)	4.95 (1.58)	.77	5.12 (1.57)	5.45 (1.23)	.37	4.94 (1.54)	4.52 (1.72)	.20
Noise	4.94 (1.69)	5.36 (1.59)	.04	5.00 (1.77)	5.22 (1.65)	.52	4.88 (1.64)	5.48 (1.53)	.03*
Clothing	3.68 (1.29)	3.74 (1.25)	.89	3.93 (1.23)	3.75 (1.19)	.17	3.46 (1.31)	3.74 (1.30)	.30
N	129	129		60	60		69	69	

Note: all scores are on 1-7 Likert scale, and all scores are recoded such that 1 is bad or low, and 7 is good or high. Significance levels are based on nonparametric analysis. Standard deviations are given in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, after multiple testing correction.

Looking at the other indoor factors, and taking male and female participants together, we do not observe lighting satisfaction to be significantly different between conditions. The same holds for the effects of light on perceived performance. Similarly, we find no difference for noise satisfaction between conditions. However, it is reported to improve performance in the hot conditions. Here also, we note that noise was kept constant between conditions. Interestingly, participants actually predict noise to improve performance compared to the control condition. We suggest that in the control condition, when the heaters only produced noise, participants perceive the noise on its own as potentially hindering performance. In the hot conditions the noise of the heaters may be driven to the background by the more salient temperature. Also, in the hot condition there is a justification for the noise. Finally, clothing satisfaction and hindrance do not differ between conditions.

5.4.3 Gender Differences and Temperature

Following recent studies of gender differences and temperature effects on performance, we examine the satisfaction measures when controlling for gender. Interestingly, the general dissatisfaction and increased hindrance of temperature are reflected in our male sample only. These findings are presented in the middle two columns of Table 5.1. Our results are in line with Chang and Kajackaite (2019), such that males dislike hot temperatures and report to suffer more from heat as compared to women. This notion is further supported by the observation that temperature experience differs between genders when related factors do not. When we compare air quality satisfaction and its hindrance between the two conditions, we find that both men and women dislike the hot temperature condition equally compared to the control condition. We note that additional (marginally) significant inconsistencies are seen for rating factors that are stable between conditions such as noise and light. Those discrepancies are correlated with the temperature manipulation (e.g., a potential demand effect; also see limitation section). [71]

Summarizing, we find that, as expected from the manipulations, temperature significantly lowers satisfaction and the perceived performance on the task, but only for the male sample. As such, as the commonly used hypothesis regarding the link between comfort and productivity predicts, we expect to find a decrease in performance on the performance measures for men, but not for women.

5.4.4 Performance Measures

Panel A of Table 5.2 shows the non-parametric results for the performance measures. We find no significant difference between control and hot conditions on any of the three performance measurements for the full sample. Only for women do we find a marginally

[71] The interaction between temperature manipulation and gender for both temperature satisfaction and temperature hinderance are both significant at $p<001$.

significant difference (T=−1.75, p=0.08; *d*=0.30) between the performance on the CRT original between the control condition (M=1.26, SD=1.09) and the hot condition (M=1.61, SD=1.24).[72] Note that performance is increasing rather than decreasing. We conclude from these first results that the temperature has no direct effect on performance for men and women on our performance measures. If anything, we find weak support in line with Chang and Kajackaite (2019), as women seem to improve rather than decrease their performance on one of the three tasks in the hot temperature condition.[73]

Table 5.2 Main Results of Performance and Risk Measures

				Men			Women		
	Control	Hot	*p*-value	Control	Hot	*p*-value	Control	Hot	*p*-value
Panel A. Performance Measures									
CRT original (scored 0-3)	1.67 (1.61)	1.76 (1.56)	.49	2.13 (1.07)	1.95 (1.03)	.34	1.26 (1.09)	1.61 (1.24)	.08
CRT Extended (scored 0-3)	1.53 (1.09)	1.71 (1.07)	.21	1.85 (1.04)	2.03 (1.02)	.33	1.26 (1.07)	1.42 (1.03)	.37
Heuristics Battery (scored 0-15)	6.34 (2.22)	6.26 (2.11)	.86	7.33 (2.12)	6.83 (1.98)	.18	5.48 (1.93)	5.83 (2.13)	.32
N	129	129		60	60		69	69	
Panel B. Risk Behaviour Elicitation									
Risk Elicitation (scored 1-10: 1 = extremely risk-loving, 10 = extremely risk averse)	6.11 (1.74)	5.9 (1.99)	.45	5.70 (1.85)	6.29 (2.05)	.12	6.48 (1.57)	5.61 (1.89)	.01*
N	111	113		53	51		58	62	
Panel C. Self-reported Risk Attitude									
General Risk Attitude (scored 1-10: 1 = risk-averse, 10 = fully prepared to take risk)	5.77 (1.91)	5.43 (1.75)	.12	6.08 (1.80)	5.40 (1.77)	.03*	5.49 (2.00)	5.46 (1.74)	.97
N	129	129		60	60		69	69	

Note: For all panels except C, all significance levels are based on parametric analysis. For panel C, significance levels is based on nonparametric analysis. Standard deviation are given in parentheses. * p < 0.05, ** p < 0.01, *** p < 0.001

5.4.5 Risk measures

Risk preference elicitation task. As expected from a strong body of research (for an overview, see Byrnes et al., 1999), a baseline difference in risk behavior is observed when comparing the control conditions as can be seen in Table 5.2, panel B. Based on parametric independent sample t-tests, men (M = 5.70, SD = 1.85) are significantly more risk-taking as compared to women (M = 6.48, SD =1.57; t = −2.42, p < 0.05; *d*=0.45), in line with the literature.

[72] For post-hoc effect size sensitivity analysis, see Appendix Table 5.9

[73] The results do show a clear and significant difference in CRT performance between genders. These results are in line with earlier findings (Brañas-Garza et al., 2019; D. C. Zhang et al., 2016) and are suggested to be a result of gender difference in either math proficiency (for the self-reported math proficiency per gender, see Appendix Table 5.3; Welsh et al., 2013) or math self-efficacy (Brañas-Garza et al., 2019).

For the risk elicitation measure, participants in general do not differ between conditions. However, when we look at the gender subsamples, the picture changes. First, although men do not differ significantly in risk preference between conditions, women are significantly more risk loving in the hot condition ($M = 5.61$, $SD = 1.89$) compared to the control condition ($M = 6.48$, $SD = 1.57$; $t = 2.75$, $p < .01$; $d= 0.50$). As such, for women the risk and heat hypothesis appears to be a valid prediction. [74]

When comparing the risk preferences of women in the hot condition with the control condition of male risk preference, we observe that women do not only become more risk loving in a hot condition, but that their risk preference becomes equal to that of men in a normal control situation.

General risk attitude. For the general risk attitude question "Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?" (See Table 5.2, panel C), men report to be less prepared to take risk when asked in a hot condition ($Mdn = 6.5$) compared to the control condition ($Mdn = 6$; $z=2.1$, $p < .05$; $d=0.38$).[75] This is surprising, as we explicitly ask participants to reflect on their general risk attitude. This question has repeatedly shown to be stable over time and context independent, and as such, is supposed to be a stable predictor for risk behavior. Women do report a stable attitude independent of conditions.[76]

When looking at the domain-specific risk attitudes, only one differs significantly between conditions: Men predict to be less risky on work-related issues in a hot condition ($Mdn=6$) compared to the control ($Mdn=6.5$; $z =2.19$ $p=0.028$; $d=0.42$) condition.[77] For an overview of these results, see Appendix Table 5.6. This result remains significant when applying the Benjamini-Hochberg rank-dependent multiple testing correction (Benjamini & Hochberg, 1995) on the critical p-value threshold with a Q (false discovery rate) of 15%.[78]

5.5 Discussion

---

[74] The interaction is significant at $p<.01$.

[75] Note that the risk aversion scores are inverse for both measures: In the general attitude measurement, a low score equates risk aversion, whereas in the risk elicitation measure, a high score shows a late (or no) switch to the risky lottery, synonymous for risk averse behavior according to the authors of the measure.

[76] When verifying the predictive power of the general risk attitude question with the risk behavior as suggested by Falk et al. (2016), we find that in our sample the general risk attitude is not correlated with risk behavior. Moreover, we find a negative correlation in the control condition between self-reported risk attitude and risk behavior (see Appendix Table 5.6). These results do not support the validity of the self-reported risk attitude as a proxy for risk behavior. However, we only find a marginally significant interaction between temperature manipulation and gender for self-reported general risk attitude with $p=.08$.

[77] The interaction is significant at $p<.05$.

[78] McDonald (2014) claims that a Q between 10% and 20% would entail relevant results, and underline that Q should not be mistaken for a P-value. For an overview of the critical value for 15% False Discovery Rate (Q) per rank used see Appendix Table 5.7.

5.5.1 Conclusion

The increasing frequency of heatwaves, and outside temperatures that used to be exceptional, raises important questions about the impact of temperature on human performance. Of course, outdoor temperature does not need to be harmful given the mitigating effect of buildings, acting as a "shield" against temperature changes and pollution. There is evidence of a positive effect of building quality on human performance and productivity (e.g. Palacios et al., 2020). But research measuring indoor climate also shows negative performance effects resulting from exposure to adverse indoor conditions (e.g. Künn et al., 2019; Wang, 2017). Given that we spend roughly 90% of our time indoors, the effect of these adverse conditions warrants research. Understanding the effects of indoor temperature on human performance is crucial in determining and optimizing the daily indoor environment in work places and beyond.

The focus of this study is twofold: First, we assess the effect of hot temperatures on decision quality, and second, we answer the question whether peoples' stated experiences regarding these temperatures are related to this decision quality. In this study, we assessed the effect of adverse temperature by manipulation of the indoor temperature to 28° C over a two-hour period, compared to a control temperature of 22° C.

From the expectation that rational decision-making would suffer under adverse temperatures, more reliance on intuition would lead to a lower score on the Cognitive Reflection Task and to more biased responses in the Heuristic Battery. However, no significant difference on performance between the hot and control conditions were identified in this study. When looking at risk, a factor often associated with decisional quality and furthermore proposed to be correlated with the intuition-rational trade-off (Leith & Baumeister, 1996), we observe only an increase of risk preference in hot conditions for women.

Comparing these results with self-reported measures show some essential discrepancies. First, in our sample, only men find the hot condition significantly less satisfactory as compared to the control condition. Women do not seem to make a distinction between conditions. Furthermore, when asking to what extent temperature has an influence on performance, men predict that the hot temperature significantly hinders their performance. Again, women do not make this distinction.

The discrepancy between self-report and actual behavior is of crucial importance for the literature regarding the effects of indoor climate. Currently, self-reported measures are commonly used as a proxy for performance or productivity, yet this study shows that men are consistently overestimating the effect of adverse temperatures on performance. First, the discrepancy between the actual performance outcomes and the perceived hindrance from adverse temperature for men shows that men would have expected to have performed better in the control condition, which they did not. If policy makers would have assessed this self-perceived hindrance only, they might have spent significant effort and resources to improve indoor temperature conditions. In our study, however, we show that this would not result in an actual increase in performance.

On the domain of risk, we find that men assess their own daily willingness to take risk in general and in work situations to decrease when they are asked about this in the hot condition. This is surprising, since this measure is aimed at assessing the general self-reported risk preference, independent of any manipulation, and would thus be expected to be stable across conditions. For women, no significant difference between conditions is found. As for actual risk behavior, we find no difference between conditions for men.

These results have at least two implications for future indoor temperature (and indoor climate) research. First, we repeatedly find inconsistencies between the self-reported and actual effects of the indoor climate on performance. Specifically, men are overestimating the negative effect the temperature has on their performance. This shows that the use of self-reported measures as a proxy for actual performance is unreliable. Future research should focus on more direct measures of human performance and productivity than self-reported indoor climate satisfaction. Second, our research supports the recent findings of Chang and Kajackaite (2019) that gender plays a moderating part in the effect of temperature on performance. This underlines the conclusion from Kingma & Van Marken Lichtenbelt (2015) that one universal temperature standard does not fit the whole population. Gender differences have to be taken into account in any situation when we include temperature as an influential factor.

5.5.2 Limitations

Three specific limitations are worth discussing. First, a multitude of factors could mediate our results. We control for many relevant variables, yet we cannot exclude the possibility that some factors confound our results. According to Zhang, De Dear and Hancock (2019), the following factors should be considered regarding the effect of the thermal environment on performance:

Environment-related factors include intensity and duration of the indoor environment. We carefully control temperature and keep all other relevant factors constant between conditions. We include an adaption time that extends the total exposure time beyond most comparable studies. However, it is possible that higher temperatures would lead to differences in performance on the (heuristics) tasks battery (Parsons, 2014). For instance, Zhang, De Dear and Hancock (2019) found that reasoning declines from temperatures upwards of 28° C. We justify our decision for the temperature levels based on earlier research and our goal to generalize our finding to a realistic working environment of high skilled workers. By doing so, we inevitably limit the external validity of our results for higher temperatures. Finally, although we measure a multitude of variables between conditions (see Appendix Table 5.2), unobserved variables could inadvertently influences the results.

Performance-related factors include all individual factors such as age, gender, skill level, acclimation level, and emotional state. We control for individual differences between groups regarding gender, math skill, education level, age, and thermostat preference (see Appendix Table 5.4). We apply random sampling to counter unobserved variables, such as emotional state, to distort our results. The sample size is limited as the adaption (or acclimation) time required takes more resources than in comparable studies. However, we are confident that addressing the exposure time is a key advantage of our experiment relative to the current literature. Regarding participant age, the sample mainly consists of students around the age of 22 (M = 21.57, SD = 2.41). We attempted to recruit an age category representing an older population (older than 50), but recruitment turned out to be difficult. Moreover, the level of English language skills and task comprehension forced us to exclude a significant part of the successfully recruited “older” sample. The educational background of the majority of our sample (Business and Economics students) increased the likelihood of recognition of the type of tasks we assessed, and previous exposure to these constructs can influence results (we will discuss the results of multiple exposure to the CRT test below). Usage of the relatively unfamiliar extension of the CRT (Toplak et al., 2014) and an unfamiliar heuristic battery (Toplak et al., 2011) at least partially alleviates this concern.

Task-related factors include the complexity and the type of task presented to the participant. Since all participants are performing the same tasks, no confounding effect of task type and complexity is to be expected. However, a new view on the underlying mechanism of the dual process model could explain why we do not find an effect of temperature on cognitive performance using our heuristics battery. De Neys and Pennycook (2019) suggest that the deliberate system is activated only when there is a clear conflict between a heuristic reaction and a logical reaction. It is possible that the nature of our task battery elicits either an intuitive responses or a logical solution, but without a conflict between these two. The lack of conflict, according to De Neys and Pennycook, will not reveal any potential restrictions in the deliberate system because this system is not involved in the response. We deliberately test an extensive battery of well-known heuristic problems which should increase the likelihood of conflicts in which the deliberate system is active. However, we cannot fully excluded the possibility that the lack of conflict (partially) explains why we find no difference between the two groups. We encourage further research to assess both neurological measured deliberate system activation as well as the level to which these tasks present an implicit conflict between logic and intuitive response.

Second, participants likely change behavior in anticipation of the effect of the manipulation, which is unavoidable in an experiment with temperature manipulation. All participants in the manipulation conditions (e.g., the "hot" temperature condition), are instantly aware of this manipulation when entering the laboratory. To create uniformity between groups and take away emphasis on the temperature, we asked participants in all conditions to wear a provided shirt, and in both conditions the industrial heaters were on. Moreover, the indoor climate quality scale was not limited to temperature, but included other important indoor climate variables, reducing the emphasis on temperature. However, when the participants were asked to state what they thought the experiment was about, they indeed stated (in the manipulation condition) that temperature and task performance was the major aim of the experiment. In the control condition, less than 10% stated temperature to be a decisive factor (popular guesses included the influence of "clothing" or "noise" on performance).

Finally, the choice for our test battery is the outcome of a careful trade-off between practical and theoretical considerations. Research has suggested that the CRT is robust under multiple exposure (Bialek & Pennycook, 2017; Meyer et al., 2018) and consistent over time (Stagnaro et al., 2018). Recognition of the original CRT is relatively high (46% recognized at least one question, and 20% recognized all questions).[79] For the extended CRT questions, however, only 13% recognized one or more questions. The fact that we observe no difference in performance between the classic and extended CRT supports the notion that these levels of recognition and recollection of answers do not affect the results of this study.

Welsh et al. (2013) propose that the CRT merely reflects mathematical skills. In our sample we see that self-reported math skills differ significantly between genders. Women report a proficiency of 59.07 out of 100, whereas males report 67.48 out of 100 ($p < .001$). We indeed

[79] For an overview of CRT and CRT extension recognition and recollection, see Appendix Table 5.8

find that in the total sample, men outperform women in the CRT. However, this does not affect the result in the sense that we analyze the effect of temperature on performance specifically within gender. We furthermore find no interaction between math proficiency and the effect of temperature on the CRT. Nevertheless, we cannot exclude that the risk assessment is effected by the difference in math proficiency.

5.6 Appendix

Table 5.3 Sample Descriptive Statistics

					Male (43%)				Female (57%)				
	Mean (SD)	Min	Max	N	Mean (SD)	Min	Max	N	Mean (SD)	Min	Max	N	p-value
Age	21.57 (2.41)	17	31	257	21.70 (2.36)	19	31	119	21.46 (2.45)	17	31	138	0.34
Math Proficiency (1-100 scale)	62.97 (17.9)	1	100	257	67.48 (16.42)	2	100	119	59.07 (18.26)	1	88	138	0.00***
Thermostat Preference (°C, in winter)	21.91 (2.65)	12	28	235	20.94 (2.74)	12	28	106	21.58 (2.55)	12	27	129	0.02*

Note. Statistics presented are mean values and standard deviation are presented in parentheses. *p*-values results from parametric independent sample t-tests. * indicates p-value < .05, ** a p-value <.01, and *** a p-value <.001

Table 5.4 Descriptive Statistics per Condition

	Control	Hot	*p*-value
Panel A. Indoor and Outdoor Conditions			
Indoor Temperature Average (°C)	22.44	28.65	.000***
Indoor Temperature during Task (°C)	22.07	28.01	.000***
Indoor CO2 (ppm)	692.12	726.93	.722
Indoor humidity (%)	48.87	39.06	.003**
Outdoor (°C) temperature at start of the experiment	13.88	14.65	.656
Average outdoor (°C) temperature (three days average)	14.44	13.84	.785
Panel B. Individual Characteristics			
Age	21.43	21.71	.358
Math Proficiency (1-100 scale)	63.49	62.44	.639
Thermostat Preference (°C, in winter)	21.32	21.27	.890
Education Level (0-5 scale)	2.92	3.01	.580

Note. Statistics presented are mean values and standard deviation are presented in parentheses. Panel A describes the indoor and outdoor climate conditions. ppm stands for particles per million. Panel B describes the individual characteristics per condition. Thermostat Preference stated is in winter conditions. Education level in on a 0 to 5 scale, where 0 is without high school diploma, and 5 is completed masters diploma. P-values results from parametric independent sample t-tests. * p < 0.05, ** p < 0.01, *** p < 0.001

Table 5.5 Correlation Table between the Risk Attitude Measure and the Risk Behavior Measure

	Full sample			Control			Hot		
	M	SD	1	M	SD	1	M	SD	1
1. Risk Elicitation Task (Holt & Laury, 2002)	6.05	1.91		6.11	1.76		6	2.05	
2. General Risk Attitude (Dohmen et al., 2011)	5.65	1.83	-.12 [-.25, .01]	5.78	1.90	-.05 [-.24,.13]	5.52	1.77	-.20* [-.37,-.01]
N	224			111			113		

Note. The Risk Elicitation task has missing values, the summary statistics excluded all risk attitude cases that are matched to missing values for the risk task. Correlation coefficient presented is the Spearman's rho and 95% confidence interval in brackets. * p < 0.05, ** p < 0.01, *** p < 0.001

Table 5.6 Multiple Testing Correction Panel A and Panel C for 15% False Discovery Rate level

			Men		Women	
	p-value	Q = 15%	*p*-value	Q = 15%	*p*-value	Q = 15%
Panel A. Self-reported Indoor Variables Satisfaction and Hinder						
Temperature Satisfaction	.00	Sig	.00	Sig	.16	
Air Quality Satisfaction	.00	Sig	.00	Sig	.00	Sig
Light Satisfaction	.07	Sig	.56		.02	Sig
Noise Satisfaction	.18		.58		.18	
Clothing Satisfaction	.14		.02	Sig	.81	
Temperature Hinder	.00	Sig	.00	Sig	.07	Sig
Air Quality Hinder	.00	Sig	.00	Sig	.00	Sig
Light Hinder	.77		.37		.20	
Noise Hinder	.04	Sig	.52		.03	Sig
Clothing Hinder	.89		.17		.30	
Panel B. Self-reported Risk Attitude						
Driving	.35		.07		.70	
Financial Matters	.47		.88		.31	
Sports and Leisure	.58		.37		.95	
Work	.20		.02	Sig	.67	
Health	.18		.23		.49	
Others (social)	.98		.23		.30	

Note. The p-value are the result of nonparametric ranksum tests as shown in table 1. The chosen levels of False Discovery Rates (Q) are chosen given that Q=15% implies less than 1 FDR per 7 tests. Q=5% is the most conservative FDR rate, with the highest risk of False Negatives (McDonald, 2014). Applying the FDR formula (False Discovery Rate = Expected (False Positive / (False Positive + True Positive))) to the risk domain entails that the change of two significant findings amongst 7 domains would be 28.6%. We find two significant findings (in the male sample) if we correct for a FDR as low as 15%. The significance of the general risk attitude in the male sample is robust against a FDR of 12%.

Table 5.7 Critical Value for 15% False Discovery Rate (Q) per Rank Used for Multiple Testing Correction

False Discovery Rate of 15 %		
Rank	7 items	10 items
1	0,025	0,015
2	0,050	0,030
3	0,075	0,045
4	0,100	0,060
5	0,125	0,075
6	0,150	0,090
7		0,105
8		0,120
9		0,135
10		0,150

Note. The critical p-value thresholds according to the Benjamini & Hochberg (1995) are dependent on the total amount of multiple tests. According to their rank, each level of significance will be compared to their rank critical value as stated in this table. The 7 items critical value are applied to the Self-Reported Risk Attitude (Table 5.6, panel C), the 10 items critical values are applied to the Self-reported Indoor Variables Satisfaction and Hinder (Table 5.6, panel A).

Table 5.8 Overview of Recognition and Answer Remembering for the CRT Classic and CRT Extension

	Recognize Question (%)	Remember the Answer * (%)		
	Yes	Yes	No	Unsure
Panel A. CRT Original				
Lily pads	45,52	26,04	44,03	13,43
Widget problem	26,85	42,54	59,38	14,58
Bat and ball	40,86	45,30	38,46	16,24
Panel B. CRT Extended				
Class ranking	2,72	6,38	78,72	14,89
Stock market	5,45	58,33	16,67	25,00
Barrel of water	10,89	7,02	77,19	15,79
N	257			

Note. *The percentage in the remembering column is conditional on recognition. For example: For the Lily pads, of the 45.52% that recognize the questions, 44.03 % do not remember the answer.

Table 5.9 Post-Hoc Sensitivity Analysis

	N		Effect Size *d*	
	Control	Hot	Non-Parametric Mann-Whitney	Parametric T-Test
Panel A. Majority Measures				
Full Sample	129	128	0,42	0,41
Men	60	59	0,62	0,61
Women	69	69	0,58	0,56
Panel B. Risk Elicitation Task				
Full Sample	111	113	0,45	0,44
Men	53	51	0,66	0,65
Women	58	62	0,62	0,60

Note. Effect size sensitivity is reported per groupsize. The first rows apply to the majority of all presented results in the paper. Only for the risk elicitation task, the latter rows applies, due to some exclusion cases in that sample. We present for each sample-size sensitivity estimates for parametric as well as non-parametric tests.

Chapter 6

Bounded Rationality in Social Network Analysis

How do network characteristics influence the human assessment of infection risk as network spread?*

6.1 Introduction

During the COVID-19 pandemic, social distancing and interaction limitations were the best policy alternative during the absence of pharmaceutical interventions such as vaccinations (Glass et al., 2006; World Health Organization Writing Group, 2006). Competing intentions of individuals to socially interact as well as minimizing the chance of virus infection required an assessment of the risk of COVID-19 contagion (Morse et al., 2006; Stroom, Eichholtz, et al., 2021; Sun et al., 2020; Ventresca & Aleman, 2013). Yet, the social network assessment needed to predict those risks demand complex computations (Centola & Macy, 2007; Pastor-Satorras et al., 2015). Empirical studies exploring the human understanding of network spread predominately show that human perceptions often fall short of accurate mathematical models (e.g. Buckee et al., 2021; Buckee et al., 2020; Funk et al., 2015). Attempts to predict and explain these human shortcomings in turn are complicated due to the lacking understanding of the underlying mechanism in that assessment.

This paper sets out to narrow this gap and uncover how individuals perceive risk and spread through networks. Specifically, we assess whether humans' evaluation of the risk of COVID-19 spread through social networks is based on complex computations or whether it depends on (more) easily observed characteristics of these networks. We find that the perceived risk is not solely based on the objective probability of risk: easily assessable physical characteristics have stronger predictive values than the objective probability. The implications of this paper are not restricted to disease contagion. Physical characteristics could also partially predict the perceived spread in social networks such as reputation and fame in individual networks, or innovation or information in business networks.

The human assessment of social networks is as important in daily work and private life as it is complex. Social networks, social structures made up of individuals or organizations nodes, cover a wide arrange of contexts including social media networks, business networks, information circulation, knowledge networks, collaboration networks, and disease transmission (Abraham et al., 2009; Brennecke, 2020; Brennecke & Rank, 2017; Hagen et al., 2018; Pinheiro, 2011; Sun et al., 2020). The spread through a network refers to the movement of that which is being shared via such network (Easley & Kleinberg, 2010). This can vary from information, financial gains, or innovation to emotional support, reputation, or disease (Bloom et al., 2013; Burt, 2017; Christakis & Fowler, 2009; Freeman, 1978; Granovetter, 2018b; Walker et al., 1997). However, the spread through complex connected networks is not easy to grasp. Epidemiological models such as the SIS or SIR models mathematically predict the spread within social networks (Centola & Macy, 2007; Grassly & Fraser, 2008; Newman, 2002)[80]. Without the same computation powers of computers, humans are unlikely to objectively estimate the spread in a similarly objective way in their daily uncertain environments (Gigerenzer & Gaissmaier, 2015).

The mathematical models are inaccurately predicting human social network assessment as they often abstract the fact that humans are heterogenous in their perception of risks and uncertainty (Gigerenzer & Gaissmaier, 2015)[81]. In contrast to mathematical models, which are constructed based on restricted, carefully selected, and extensive data, humans function in an uncertain world full of, to them, unknowable and incalculable probabilities and uncertainties (Gigerenzer & Gaissmaier, 2015; Hafenbrädl et al., 2016; Simon, 1956). The inability to match mathematical models' outcomes falls under the reasoning of bounded rationality: people are rationally motivated to optimize the problem, but limited mental capacity hampers or prevents them (Gigerenzer & Selten, 2002; Simon, 1990). Relatively simple rules of thumb, or heuristics, help overcome these mental limitations (Tversky & Kahneman, 1974). For risk and uncertainty,

[80] For an excellent overview, we would like to refer to Chapter 18 in the Oxford Handbook of the Economics of networks (2016), written by Lamberson, and the article by Pastor-Satorras (2015).

[81] Risk perception relates to network assessment in context of disease spread or contagion. For simplicity reasons, we equate risk with the likelihood of spread through networks. In the context of disease, the risk of contagion is synonymous with the likelihood of (general social network) spread.

these subconsciously applied shortcuts enable the construction of guestimations, or gut-feeling approximations (Gigerenzer & Selten, 2002; Kahneman & Frederick, 2005). The downside of these heuristics is that by mitigating uncertainty, limited knowledge, and computational capacity constraint, they sacrifice accuracy (Gigerenzer et al., 2011; Kahneman et al., 1982). As a results, human perception of risk often deviates in a predictive way (Siegrist & Árvai, 2020). Since increasing complexity of a network makes the spread practically incalculable, it's feasible that heuristics also subconsciously substitute the mental computations during social network assessment. Network physical characteristics, such as closeness to the source, size, and shape, could be used to inform about spread. Thus, in case rational optimization is unfeasible, heuristics provide a second-best (Gigerenzer et al., 2011). In that sense, bounded rationality might result in predictably irrational assessment errors, even when they are driven by rational motives (Sent, 2018).

The accumulation of multiple uncertainties of many real-life networks complicates the isolation and exploration of heuristics in human social network assessment. For example, it is generally not apparent which conditions need to be met in order to spread that-which-is-being-shared-through-the-network through the network. What makes a node susceptible for spread? Although much domain-specific research revolves around this pivotal question, for each new context, there is no generic set of conditions that constitute spread. For instance, innovation might not spread from one firm to another by a simple connection or tie (e.g. Amabile, 1988; Ventresca & Aleman, 2013). It might be that endured collaboration is needed, or rivalry, or a threshold similarity. Hence, a real-life connection does not necessarily equal a susceptible connection that has the potential to spread.

Second, the spread is generally heavily dependent on the a priori unknown pass-on likelihood. Given that a connection is indeed susceptible to spread between two nodes, it is not guaranteed that that-which-is-being-shared-through-the-network will pass on from one node through that connection. Thus, the spread likelihood, or pass-on rate, needs to be estimated. Mathematical models designed to estimate these conditional spreads are trained on past events. For a human beholder of a network, however, the spread characteristics of each new network are highly complex and often obscured. Both taken together, heuristics applied to social network assessment research are influenced by beliefs and assumptions on how humans deal with these uncertainties first. As a result, applied research on for instance the effect relatively simple network structure characteristics might have on perceived spread in real-life social networks, remains largely underinvestigated.

The COVID-19 pandemic has introduced a novel highly relevant real-life context for social network assessment and risk perception research. Minimizing the spread of the highly contagious and potentially life-threatening virus has been at the forefront of most international policies (World Health Organization, 2020). Nevertheless, these policies demanded assessments and evaluations of each individual to estimate what was safe behavior (Huremović, 2019). At some point, infections from social networks were not completely avoidable, and risks need to be assessed. Prior research focuses on COVID-19 risk perception. The findings often conclude that a higher COVID-19 risk perception leads to more precautionary behavior (Bruine de Bruin & Bennett, 2020; Dryhurst et al., 2020; Nelson et al., 2020; Stroom, Eichholtz, et al., 2021; Plohl & Musil, 2021; Savadori & Lauriola, 2021; Schneider et al., 2021; Sobkow et al., 2020; d'Andrea et al., 2022). A large strand of the literature is devoted to the possible determinants of Covid risk perception such as political inclinations (e.g. Bruine de Bruin et al., 2020; Shao & Hao, 2020), information and media sources (e.g. He et al., 2021; Karasneh et al., 2021; Liu et al., 2020; Stroom et al., 2021), and personal attitudes and emotional states (e.g. Dryhurst et al., 2020; Mertens et al., 2020; Qian & Li, 2020; Zhong et al., 2021). Social network assessment has only been mentioned to optimize the distance strategy(Block et al., 2020). How characteristics of the network itself influenced decision-making, has been largely overlooked.

The emphasis on virus contagion qualifies COVID-19 as the fitting setting to investigate how the general population process spread through social networks. Moreover, the abundance of information provided mitigates many aforementioned concerns. First, every person could be considered susceptible. Although not all infected people would get sick or even show symptoms, to this day, no clear exception of general susceptibility is known. Second, the risk of infection was mainly communicated through a reproduction-metric (R): the average amount of secondary infections that follow from one primary infection. Although multiple factors influence this risk, the R metric gave a universally understandable indication of risk in social networks. Consequently, taking these uncertainties out of the equation, other factors influencing human social network assessment can be isolated and identified in a real-life context.

In this study we explore how people perceive risk or spread in social networks. We specifically focus on whether the physical characteristics, such as size and shape, of a network are aiding otherwise complex computations. We question whether individuals perceive some network structures as riskier than others. Using a best-worst choice experiment, we let individuals repeatedly rank three varying infected networks, randomly drawn from a set of fourteen, from most risky to least risky. We find that the perceived risk is not solely based on the objective probability of risk, since easily assessable factors have stronger predictive values than the objective probability. Furthermore, we show that we are able to relatively accurately predict the observed network preferences based on the preference estimates associated with the network characteristics. Our results consolidate that humans' processing of risk in networks is not completely rational and also depends on the simple characteristics of these networks. Heuristics have strong correlational power with objective risk in isolation, yet remain to resonate and dominate in complex networks even when other factors are also detrimental to the objective risk. The often-complex mental calculation of objective risk dispersion in networks seems at least partially substituted by a heuristics-driven approach.

6.2 Methods

6.2.1 Best-worst choice experiment

We investigated individuals' perception of the risk to connect to infected networks using a best-worst choice experiment. In this experiment, participants were presented with eight sets of three networks and for each set they had to choose the 'best' and 'worst' network according to the highest and lowest risk of infection for themselves, respectively. We surveyed 1,002 Dutch individuals via the Flycatcher panel. Flycatcher is an academically oriented Dutch research organization that established a high-quality panel representing the Dutch population[82]. Flycatcher randomly sampled our participants from their extensive panel for which participation was reimbursed. This research was reviewed and approved by Maastricht University's Ethical Review Committee Inner City Faculties (ERCIC_195_09_06_2020).[83]

[82] For comparison of our sample with (approximations) of the Dutch population characteristics, see Appendix Table 6.8

[83] The experiment was done in combination with a short survey on working from home for chapter 4. None of the questions before touched the issue of risk or infections. We therefore regard the data as independent and analyze them in isolation.

The experiment consisted of three parts. First, participants were introduced to the experimental setup and provided with the essential information and parameters that govern the experiment throughout. Participants were shown an example network and were trained to identify themselves in the network (green icon), an infected person at the start of the network (red icon), and the remaining people in the network (blue icons; see Appendix Figure 6.6). Each person had contact with another person in the network in a specified order from right to left. We will refer to each individual 'person' in the network as a node. Participants were informed that the network connections were directional in the sense that there was a specified notion of flow and that all paths were one-way following the directional arrows. For each connection indicated by an arrow, the likelihood that the infection was passed to the next node was 50%. The contagion was uncontested, meaning that more links did not decrease the change of infection of each link (Centola & Macy, 2007). The respondents' task associated with each set of three networks was then explained. When connecting to the networks at the indicated point (a dotted directional arrow), respondents had to rank the network with the highest risk of infection for themselves at the top and the network with the lowest risk at the bottom. In each network, there was at least one person infected. We defined risk as the likelihood that the respondents themselves would get infected. The training section concluded with the following two questions to assess the level of understanding by each respondent regarding the information provided: "How likely is it to get infected when you connect directly to an infected node?" and "How are you supposed to rank the network riskiness?" (see Appendix Figure 6.7).

Second, we presented the actual best-worst exercise for infection probability assessment. Specifically, we assigned eight randomly selected subsets of three networks drawn from the total network pool to each participant, where the subsets were also random across participants. For each of the networks, participants were reminded to rank them on the level of perceived risk (i.e. infection probability) of the green node: the network for which a participant perceived the infection probability as highest would rank number one, and the network for which the infection probability was considered lowest would rank in third place (see Appendix Figure 6.8). Note that we give participants full information about the individual infection probability and the network structure. A person who is experienced in combinatorics could calculate the infection probabilities. The real-life complexity, however, does not deem it likely that these probabilities are generally calculated. It is also important to realize that the willingness to take risk should not influence the assessment of risk. In the context of the pandemic, one might argue that the risk is heterogeneously weighed against the reward of social interaction.

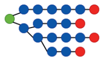
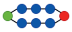
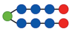
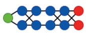
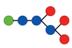
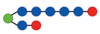
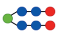
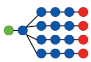
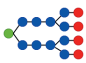
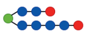
Therefore, we don't survey absolute willingness to connect to each network, but ask participants to rank subset relatively.

The final part of the experiment consisted of gathering additional information for further analysis. Besides demographic questions, we asked participants to rate their willingness to take risk in general, related to health, and related to trust in others, on a scale from 0 (not at all willing to take risks) to 10 (very much willing to take risks). Furthermore, we connected our participants to COVID-19 exposure based on their postcode. At this postcode level, we identified the number of infections, hospitalizations, and deaths related to COVID-19 at the time of the survey.

6.2.2 Network construction

For the experiment, we considered 14 different networks that varied on a multitude of characteristics. For each network, Table 6.1 shows the identifier, an icon showing the structure of the network as displayed in the survey, the relevant characteristics (not explicitly given to the participants), the total number of views or occurrences in the subsets presented to the respondents and the rank frequencies based on the full sample (see Section 2.4).

Table 6.1 Characteristics and rank frequencies of the 14 networks (based on the full sample)

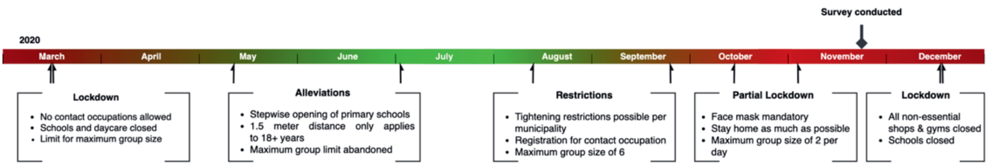
		Network Characteristics					Viewed	Ratio Ranking		
Network	Figure	Infection Probability	Percentage Infected	Number of Direct Links	Shortest Path	Number of Paths	Total	First	Second	Third
1		8.78%	21.05%	2	5	4	1234	22.93%	28.28%	48.78%
2		12.11%	14.29%	2	4	2	1310	12.75%	39.31%	47.94%
3		12.11%	25.00%	2	4	2	1425	18.67%	46.95%	34.39%
4		6.25%	25.00%	1	4	1	1443	13.86%	26.47%	59.67%
5		12.50%	33.33%	1	3	1	1348	41.47%	32.42%	26.11%
6		14.27%	20.00%	2	5	6	1329	27.77%	44.17%	28.07%
7		23.57%	9.09%	2	5	10	1189	24.98%	19.76%	55.26%
8		26.17%	20.00%	2	5	16	1221	43.16%	35.63%	21.21%
9		17.19%	50.00%	1	3	3	1100	68.91%	23.36%	7.73%
10		26.17%	25.00%	2	2	2	1096	63.14%	23.54%	13.32%
11		23.44%	33.33%	2	3	2	1078	52.97%	36.36%	10.67%
12		11.38%	23.53%	1	5	4	1038	28.71%	30.83%	40.46%
13		10.64%	28.57%	2	5	4	961	26.74%	36.52%	36.73%
14		15.23%	25.00%	2	3	2	956	34.73%	40.69%	24.58%

The networks in Table 6.1varied on 1) the number of direct links to the participant, or the number of connections to the green node, which is either 1 or 2; 2) the total number of paths leading from any infected node to the green node, and varying between 1 and 16; 3) the shortest pathway from any infected node to the green node, ranging from 2 to 5; 4) the actual probability of infection of the green node, which ranges between 6.25% and 26.17%; and 5) the number of infected nodes relative to the total number of nodes, or the percentage of infections per network, ranging between 9.09% and 50%. For all networks applies that they contain only one green node but can contain multiple red nodes, an infected red node never connects directly to a green node, and all blue nodes can connect to or receive from multiple nodes.

6.2.3 Data collection period

The data collection took place in November 2020. At this time, the Netherlands had been in a (partial) lockdown for over 8 months due to the COVID-19 pandemic, since March 2020. Figure 6.1 shows that the first severe restrictions were swiftly introduced as a reaction to the first cases of the virus in the Netherlands. After 3 months of hard lockdown, the government lifted most restrictions in May through July, leading to a relatively optimistic summer. Between August 6 and 18, the second wave of the virus started with tightening restrictions yet again. From August through November, four distinct, consecutive, and increasingly more severe national restrictions were introduced. At the time of the data collection, staying home as much as possible was the leading recommendation.[84] Participants were confronted with a severely limited social life (a maximum of two social contacts simultaneously, and no access to social events or venues).

Figure 6.1 Survey timeline



6.2.4 Sample characteristics

Our initial sample drawn from the Flycatcher panel consisted of 1,002 Dutch participants. For this initial sample, no demographical exclusion criteria were applied. Correct answers to the two questions raised at the end of the training section to assess whether

[84] It is important to note that the policy recommendation to stay home as much as possible in the Netherlands is not the same as a lockdown. The Netherlands has had a tendency to have policies without strongly enforced regulations but leaving the final decision to each individual. The Dutch population was effectively expected to assess the validity of their behavior on their own. This increases the realistic nature of our experiment presented to the Dutch participants. For instance, chapter 2 show that individuals' assessments differed across individuals, and under different Dutch recommendations.

participants understood the information provided were crucial for properly assessing the risk between networks. Therefore, we excluded 305 participants who did not provide the correct answers to both questions. As a result, 697 participants remained in our final sample for further analysis.

Table 6.2 compares the included and excluded participants statistically. Specifically older participants (difference of 5.92 years; $p<.0001$) and level of education (62.4% of the included participants were highly educated versus 36.7% of the excluded participants; $p<.0001$) play an important role in understanding and replicating the basic assumptions of the network questions. These differences are not surprising, as the nature of the task can be perceived as technical and hypothetical. Abstract thinking is required so that personal characteristics related to cognitive performance can predict a failure to meet the basic understanding of the choice tasks. Importantly, we do not see a difference in risk preferences, again solidifying that the main driver for exclusion might be related to cognitive performance or being able to perform high-level abstract thinking. In our final sample, 56.0% of the participants were male, with an average age of 42.09 years (SD=12.43), and 62.4% completed higher education.

Table 6.2 Summary statistics for the included and excluded participants

	Included (N=697)	Excluded (N=305)	Total (N=1002)	p-value
Gender				0.039 (1)
Male	390 (56.0%)	192 (63.0%)	582 (58.1%)	
Female	307 (44.0%)	113 (37.0%)	420 (41.9%)	
Age in years				< 0.001 (2)
Mean (SD)	42.09 (12.43)	48.01 (11.81)	43.89 (12.54)	
Level of education				< 0.001 (1)
Low	31 (4.4%)	34 (11.1%)	65 (6.5%)	
Medium	231 (33.1%)	159 (52.1%)	390 (38.9%)	
High	435 (62.4%)	112 (36.7%)	547 (54.6%)	
Willingness to take risk in general				0.718 (2)
Mean (SD)	5.26 (2.21)	5.32 (2.60)	5.28 (2.33)	
Willingness to take risk with health				0.262 (2)
Mean (SD)	4.39 (2.08)	4.65 (2.54)	4.47 (2.23)	
Willingness to take risk in social context				0.839 (2)
Mean (SD)	5.61 (2.12)	5.54 (2.35)	5.59 (2.19)	

Note: 1) Pearson's Chi-squared test, 2) Kruskal-Wallis rank sum test

6.3 Data analysis

6.3.1 Modelling approach

To model the rankings of the $J = 3$ networks in each of the $S = 8$ choice sets from $N = 697$ respondents, we used a panel mixed logit (PML) model that analyses the rankings as best-worst choice data. This means that the best or first ranked network profile in each choice set is coded as 1, the least or third ranked profile as -1 and the middle or second ranked profile as 0. This modelling approach relies on random utility theory and is also referred to as maximum difference (MaxDiff) modelling (Marley & Louviere, 2005). It has been used in many applied economics settings; see, e.g., Mühlbacher et al. (2016) for analysing health(care) choices and Yeh et al. (2020) for analysing food choices. Note that strictly speaking we do not measure preferences here but beliefs. Nevertheless, we consider a utility-based approach appropriate as participants were voluntary and intrinsically motivated to provide true answers to the ranking questions. Hence choice utility and beliefs are aligned.

The MaxDiff PML model starts from McFadden's (1973) random utility formula that defines the utility that respondent n, $n = 1, \ldots, N$ attaches to network profile j, $j = 1, \ldots, J$ in choice set s, $s = 1, \ldots, S$, as the sum of a systematic and a stochastic component:

$$U_{njs} = V_{njs} + \varepsilon_{njs} = \boldsymbol{x}_{njs}\boldsymbol{\beta}_n + \varepsilon_{njs}, \tag{1}$$

where $\boldsymbol{x}_{njs}$ is a k-dimensional vector containing the attribute levels x_{njsk} of network profile j in choice set s for respondent n and $\boldsymbol{\beta}_n$ is a k-dimensional vector of preference parameter values β_{nk} representing the individual-specific effects of the attribute levels on the utility. The stochastic component ε_{njs} is the idiosyncratic error term, which is assumed to be independent and identically Gumbel distributed across n, j, s.

In our analysis, we extended the systematic component of the utility to account for preference heterogeneity by adding interactions between attributes and individual-specific covariates as follows:

$$V_{njs} = \boldsymbol{x}_{njs}\boldsymbol{\beta}_n + \sum_{d=1}^{D_c} \beta_{nkd}\delta_d(c_{nk})x_{njsk}\,, \tag{2}$$

where $\delta_d(c_{nk})$ represents one covariate value from a total of D_c possible values for respondent n. Also, we used coded levels between -1 and 1 for both continuous and categorical attributes so that all attributes are on the same scale and the attribute effects can easily be compared with each other. For continuous attributes this is realized by subtracting the midpoint of the interval $[x_{\min,k}, x_{\max,k}]$ from each value x_{njsk} and dividing the result by half of the difference between $x_{\min,k}$ and $x_{\max,k}$. For categorical attributes, we used effects-type coding which codes the relevant category as 1, the reference category as -1 and all other categories as zero.

In our ranking experiment, an individual n ranks network j in choice set s first and network j' third, or considers network j as most important and network j' as least important. The utility difference between U_{njs} and $U_{nj's}$ is then the greatest among all $J(J-1) = 3(3-1) = 6$ possible utility differences involving pairs of network profiles a person could choose from the choice set.

Following Louviere et al. (2015), the choice probability *P* of individual *n* ranking profile *j* first and profile *j'* third in choice set *s* can be written as:

$$P_{njj's} = \frac{\exp\left(V_{njs} - V_{nj's}\right)}{\sum_{i=1}^{J} \sum_{i'=1,i'\neq i}^{J} \exp(V_{nis} - V_{ni's})} \,. \tag{3}$$

We estimated MaxDiff PML model (3) using Hierarchical Bayesian estimation in the JMP 16 pro MaxDiff Platform (based on 10,000 iterations, with the last 5000 used for estimation). We assumed normally distributed preference parameters without correlation between attributes. These random parameters accommodate unobserved heterogeneity in the respondent preferences. The mean utility function is thereby the sum of the mean attribute effects (Train, 2009).

6.3.2 Cluster analysis approach

We first estimated a MaxDiff PML model for the entire sample and then investigated the heterogeneity in the individual utility estimates by comparing the subject standard deviations to the mean attribute effects. These subject standard deviations were of the same size or even larger than the mean estimates, indicating the need to identify respondent segments. We, therefore, clustered the individual utility estimates from the PML model using Ward's hierarchical cluster analysis and estimated a full interaction effects PML model with the cluster as a covariate using utility formula (2). Clusters were identified based on demographic information, including gender, age, and risk preference. This final PML analysis allows revealing differing and even opposing preferences between clusters.

6.4 Results

6.4.1 Rank frequencies of the networks

Table 6.1 shows the rank frequencies for each network overall choice tasks evaluated by the 697 participants. Considering all subsets that included a specific network, it shows how often that network has been ranked first, second, or third on riskiness. From these simple rankings, we observe that network 9 is most often perceived as most risky from its subset and least often as least risky network. Network 4 is relatively most often scored as least risky, but this does not automatically imply that it is also the least often considered as most risky. In fact, network 2 is least often ranked first (12.75% versus 13.86%). We can use these rankings to compare the overall trends with the infection probability, i.e. the actual percentage risk of infection when connecting to a network. Although network 4 is correctly perceived as the least risky compared to all networks, network 9 is by far not the riskiest network to connect with. Specifically, 4 out of the 14 networks have a higher infection probability than network 9. These results suggest that, in contrast to the rational point of view, other characteristics beyond the objective probability influence the rankings made by participants.[85]

6.4.2 Main Panel Mixed Logit model estimates

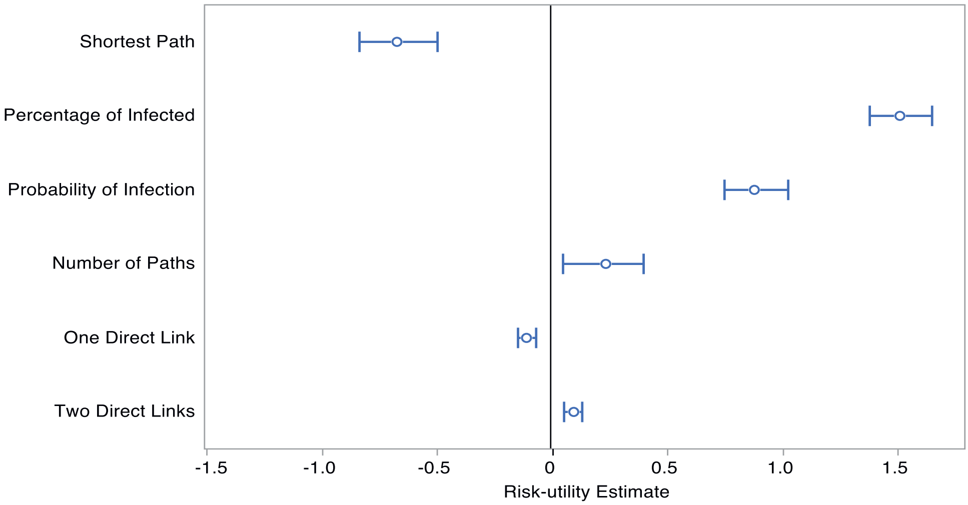
Given the need for further clarification on the rankings made, we estimated a MaxDiff PML model to obtain the random utility estimates associated with the five network characteristics included in Table 2. Figure 6.2 shows the main risk-utility estimates for each characteristic that has been rescaled to lie between -1 (minimum) and 1 (maximum). A significant positive estimate for a characteristic implies an increase in the perceived level of riskiness of the network when moving from 0 (midpoint) to 1 for that characteristic. A significant negative estimate suggests that such increase in the characteristic will decrease the perceived riskiness of a network. In general, large estimates (in magnitude) signal high-risk characteristics, whereas small estimates indicate low-risk characteristics (see Appendix Table 6.9, Panel A for the estimates).

The main risk-utility estimates reveal that the percentage of infected nodes in a network is the most important determinant of risk assessment. By increasing this percentage, the perceived risk-utility gets larger (coefficient of 1.52, 95% CI [1.39; 1.66]). The objective probability of infection has a surprisingly weaker effect, increasing perceived risk by only 0.89 (95% CI [0.76; 1.03]). The length of the shortest path additionally has a strong effect in the sense that the closer the participant is to an infected node, the higher the perceived risk (risk-utility coefficient of -0.67, 95% CI [-0.83; -0.49]). Each additional path from an infected node to the participant (number of paths) increases the perceived risk by 0.24 (95% CI [0.06; 0.41]). Finally, regarding the number of direct links, having two direct links to the participant increases the risk-utility by 0.10 (95% CI [0.06; 0.14]), whereas one direct link decreases it with the same value

[85] In section 3.4, we further convert these ratings into a single metric which allows us to compare the ranking between networks.

(due to the effects-type coding). To summarize, networks with an infected node close to the participant, a relatively large number of infected nodes, a high probability of infection, and two direct links, are perceived as riskier compared to alternatives. It is noteworthy that the attributes more with higher risk signaling estimates (e.g. shortest path and percentage of infections) are also two attributes of which the information is often available in real life settings.

Figure 6.2 Main model estimates



Note: the lines indicate the 95% confidence interval

6.4.3 Cluster Panel Mixed Logit model estimates

The individual-level utility estimates generated by the PML model allowed us to identify clusters of respondents with similar risk preferences. Ward's hierarchal cluster analysis identified three distinct clusters in our sample (largest cubic clustering criterion of 73.5%; Appendix Figure 6.9 shows the 2-dimensional constellation or scatter plot displaying the cluster memberships of the individual respondents). Table 6.3 compares the demographic information and risk attitude of the clusters. The main differences can be found in gender and education. Cluster 1 (N = 293) contains more females and relatively more participants with lower levels of education. Cluster 2 (N = 219) is dominated by male participants and high levels of education. Cluster 3 (N = 185) falls in the middle category with respect to gender and education. The characteristics, age and willingness to take risk, are not significantly different between clusters, although we see some (insignificant) trend for more risk taking in Clusters 2 and 3.

We must note that we describe the clusters informatively, without clear pre-constructed motives. The clustering approach aims primarily at improving the PML model fit. Nevertheless, the clusters are distinct enough in their characteristics to explore potential heterogeneity in the population regarding risk assessment in social networks.

Table 6.3 Summary Statistics

	Cluster 1 (N=293)	Cluster 2 (N=219)	Cluster 3 (N=185)	Total (N=697)	p-value
Gender					0.005 (1)
Male	143 (48.8%)	135 (61.6%)	112 (60.5%)	390 (56.0%)	
Female	150 (51.2%)	84 (38.4%)	73 (39.5%)	307 (44.0%)	
Age in years					0.370 (2)
Median	41.00	40.00	42.00	41.00	
Mean (SD)	41.91 (12.14)	41.47 (12.74)	43.12 (12.53)	42.09 (12.43)	
Level of education					0.025 (1)
Low /Medium	124 (42.3%)	67 (30.6%)	71 (38.4%)	262 (37.6%)	
High	169 (57.7%)	152 (69.4%)	114 (61.6%)	435 (62.4%)	
Willingness to take risk in general					0.221 (2)
Median	5.00	6.00	4.00	5.00	
Mean (SD)	5.09 (2.17)	5.41 (2.14)	5.37 (2.34)	5.26 (2.21)	
Willingness to take risk with health					0.235 (2)
Median	4.00	4.00	4.00	4.00	
Mean (SD)	4.24 (2.08)	4.55 (2.10)	4.45 (2.06)	4.39 (2.08)	
Willingness to take risk in social context					0.694 (2)
Median	6.00	6.00	6.00	6.00	
Mean (SD)	5.52 (2.16)	5.68 (2.11)	5.69 (2.08)	5.61 (2.12)	

Note: 1) Pearson's Chi-squared test, 2) Kruskal-Wallis rank sum test

Figure 6.3 show the marginal risk-utility estimates of the characteristics for the three clusters (see appendix Table 6.9, Panel B for the estimates). Similarly to Figure 6.2, a positive (negative) risk-utility estimate of a (coded) characteristic indicates an increased (decreased) perceived level of risk when increasing the characteristic from 0 (midpoint) to 1. Compared with the main model estimates of Figure 6.2, the cluster estimates reveal significant differences in most of the attributes. The differences per cluster are specifically profound for the length of the shortest path, the probability of infection, and percentage of infected nodes. The estimates of the number of direct links do not differ significantly per the clusters, thus the plotted estimates in Figure 6.3 per cluster are equal to each other, and to the main effect estimates of Figure 6.2 (risk-utility coefficient of 0.10, 95% CI [0.06; 0.14]).

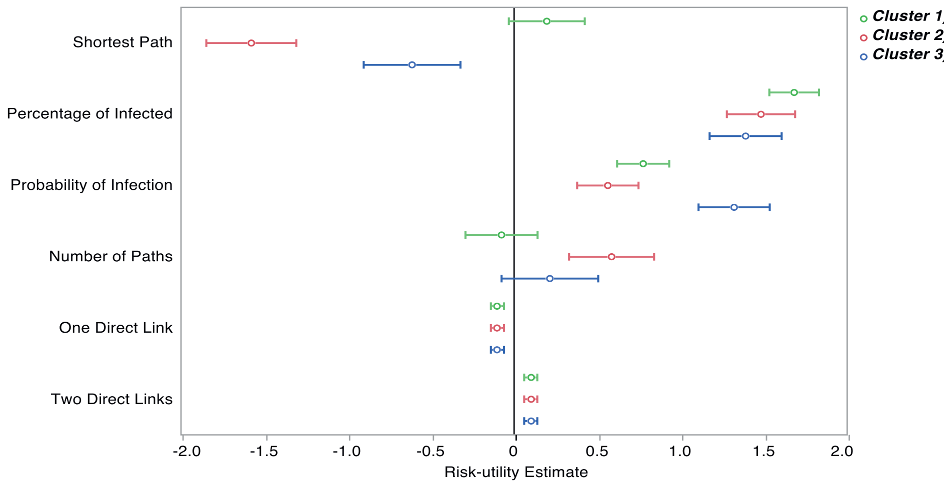
In general, the cluster 1 risk assessments are mostly driven by the percentage of infected. For this cluster, the percentage of infected has the highest risk-utility estimate out of all clusters (coefficient of 1.68, 95% CI [1.53;1.83]). The second driver, objective probability of infection, falls behind by approximately half of that impact with a risk-utility estimate of 0.78 (95% CI [0.62;0.93]). The length of the shortest path and the number of paths to an infected node are not significant drivers of risk-utility for this cluster.

Cluster 2 (higher educated and more male) is the only cluster that remains sensitive for variation in all characteristics. This cluster is particularly sensitive to changes in the shortest path, showing a marginal utility drop of -1.58 (95% CI [-1.85;-1.31]) when increasing the infection distance from an artificial 3.5 nodes (midpoint) to the maximum of 5 nodes. Also, as opposed to the other clusters, cluster 2 participants attach significant importance to the number of paths (risk-utility coefficient of 0.59, 95% CI [0.33;0.84]). For both the percentage of infected and probability of infection, cluster 2 is similar to cluster 1 (risk-utility coefficients of 1.48, 95% CI [1.28;1.69], and 0.56, 95% CI [0.38;0.75], respectively).

The risk-utility for cluster 3 participants is driven by three characteristics: the shortest path, the percentage of infections, and probability of infection. Out of the three clusters, this cluster attaches the most importance to the objective infection probability (risk-utility coefficients of 1.32, 95% CI [1.11;1.53]). Only the percentage of infected nodes has a (marginally) higher coefficient, at 1.39 (95% CI [1.17;1.61]). Finally, the shortest path to an infected node is significant with a the risk-utility of -0.61 (95% CI [-0.90;-0.32]). Similar to cluster 1, the number of paths no longer significantly predicts the risk evaluation for cluster 3.

These cluster differences reveal interesting findings. First, for all clusters, the percentage of infected nodes has a more profound effect than the probability of infection itself. For cluster 1 and 2 this effect is two to three times the effect of the objective probability. The notion that our sample consists of at least one cluster of merely rational agents who solely focus on the objective probability clearly does not hold. Second, there might exist a relation between education and adaption strategies. Ordering the clusters on proportion of higher educated participants shows a linear relation with the number of characteristics that significantly influence the risk-utility. The higher educated cluster 2 seems to base their risk preference on all characteristics, and is most sensitive to a relatively robust shortcut characteristic: closest to an infected node. Cluster 3 takes four characteristics into consideration, and lower-educated cluster 1 only takes three out of five characteristics into consideration. Finally, education alone does not to predict accurate risk assessment. Cluster 2 is the least sensitive to the objective probability of infection. This relationship is however not linear, as cluster 3 is moderate in level of education and most sensitive to the objective probability of infection.

Figure 6.3 Cluster model estimates



Note: the lines indicate the 95% confidence interval

6.4.4 PML utility-based network ranking

After identifying the driving factors for risk assessment of networks, we investigated how accurately we can predict the perceived risk of a network based on the risk-utility estimates of the specific characteristics of the network. Using equation (1) on the main model estimates as well as cluster model estimates, we computed the total risk-utility for each network and sample (the full sample and clusters), representing an overall utility-based risk score per network and sample. We ranked the utility-based risk scores per sample such that a model-based ranking of the existing networks emerged. Table 6.4 shows all 14 networks ranked per group on the overall utility-based risk scores. Networks displayed at rank 1 are predicted to be most risky based on their characteristics and how each respective (sub)sample derives 'risk utility' from them. Consequently, networks displayed at rank 14 are predicted to be deemed least risky based on that sub(group) utilities. In the last row, the networks are ranked based on the objective probability of infection. The actual risk scores per network and sample as well as the corresponding network IDs appear in Table 6.10 in the Appendix.

Table 6.4 PML utility-based ranking of the 14 networks, from most to least risky

Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Full Sample														
ID	9	10	11	8	5	14	3	13	7	6	12	4	2	1
Cluster 1														
ID	9	11	10	8	5	14	13	3	6	12	7	1	4	2
Cluster 2														
ID	9	10	11	5	14	8	3	4	13	2	7	6	12	1
Cluster 3														
ID	10	9	11	8	14	5	7	3	13	6	2	12	4	1
Objective Probability	*	*							*	*				
ID	10	8	7	11	9	14	6	5	3	2	12	13	1	4

Note: For the row 'Objective Probability', * indicates that the two adjacent networks have an identical risk probability and thus share their respective rank position.

The rankings in Table 6.4 differ across the full sample and the clusters. The most important focus is the difference between the full sample and cluster 1 rankings. We mentioned previously that the estimates of the full sample were affected by the inclusion of the cluster 2 and 3 choices. The network rankings of cluster 1 are noticeably different from those of the full sample. For example, the network ranked 8th in the full sample falls back to rank 13 in cluster 1 (network ID 2). Similarly, the networks in ranks 10 and 14 are reversed (network IDs 4 and 7). Table 6.4 shows the correlations between the full sample, clusters, and objective probability rank.

The rankings in Table 6.4 differ across the full sample and the clusters. Each PML prediction ranking is different from the other. For example, the network ranked 12th (Network ID 4) in the full sample is expected to be less risky for clusters 1 and 3, but 4 ranks riskier for cluster 2. Table 6.5 shows the correlations between the full sample, clusters, and objective probability ranks of Table 6.4. We observe that the predicted PML ranking of the full sample correlates with the objective risk ranking (rho = .74, p<.01). When the sample is divided into clusters, an order is revealed in which cluster 2 is moderately correlated with the objective rank (rho = 56, p=.04), cluster 1 more so (rho = .63, p=.02), and cluster 3 strongly (rho = .73, p<.001). These correlations suggest that the consistent application of characteristics preference and weighing derived from cluster 3 participants leads to the most accurately approximation of objective risk for our networks, whereas the consistent application of characteristics preference and weighing derived from cluster 3 participants leads to the least accurately approximation of objective risk for our networks.

Table 6.5 Correlation matrix of PML utility-based and objective probability ranking

	Objective Probability	Full Sample	Cluster 1	Cluster 2	Cluster 3
Objective Probability	1				
Full Sample	0.74**	1			
Cluster 1	0.63*	0.96***	1		
Cluster 2	0.56*	0.90***	0.82***	1	
Cluster 3	0.86***	0.96***	0.88***	0.87***	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We can test the validity or predictive power of our estimates by examining how well the model-based rankings correlate with the observed rankings for each sample. In order to compare these two rankings, we first convert the observed rank frequencies into a single metric. As shown in Table 6.1, we identified the frequencies that a network has been ranked first, second or third in all subsets presented to the participants. Ranking the 14 networks by only one of these three frequencies proves challenging, as a low score on rank 1 does not automatically indicate a high score on rank 3 and vice versa. Therefore, for each network *j* and sample *l*, the first, second, and third subset rank frequencies *RF* are converted into a single metric, which we call the observed subset mean rank *SMR*, by using them as weights for the ranks:

$$SMR_{jl} = RF_{jl}^{First} + 2 * RF_{jl}^{Second} + 3 * RF_{jl}^{Third} \tag{4}$$

By weighing the ranks with their frequencies, this metric is an improvement over a simple first or last frequency ranking. However, this does not completely alleviate the possible influence of unbalanced frequencies over the ranks. To provide more insight, we define the unanimity of the rank frequencies *RF*, which serves as a measure for the spread in the rank frequencies:

$$Unanimity(RF_{jl}) = \frac{\left((RF_{jl}^{First} - \frac{1}{3})^2 + (RF_{jl}^{Second} - \frac{1}{3})^2 + (RF_{jl}^{Third} - \frac{1}{3})^2\right)}{2} \tag{5}$$

A high unanimity suggests preference homogeneity in the frequency scoring, whereas a low unanimity suggests preference heterogeneity. A perfectly unanimous preference, meaning all participants rank a network in the same position, results in a network ranking unanimity score (eq. 5) of 0.33. A perfectly divided preference, meaning 33% of the respondents rank a network first, 33% rank it second, and 33% rank it third, results in an unanimity level of practically zero. Low unanimity can also be interpreted as the absence of a strong or dominant preference in the sample.

Defining single ranking and unanimity metrics for each network enables us to explore the network preference trends. For example, Figure 6.4 visually ranks the networks using the subset mean rank (SMR) from most risky to least risky for the full sample. Figure 6.4A shows clearly that the progress through the ranks from perceived riskiest to least risky does not perfectly coincide with a decline of objective infection probability. If the preferred ranking would be based solely and accurately on objective risk probability, the objective probability would decline with each decrease in rank. However, Figure 6.4A shows that the perceived riskiest network is not in the top five of objectively riskiest networks. Moreover, multiple networks perceived as less risky are objectively more probable to lead to infection. Figure 6.4B shows the trend of unanimity with regard to the preference-based SMR ranking. The extremes (most risky and least risky) are ranked with higher unanimity compared to the middle rankings. This implies that our full sample generally agrees more on the perceived top and bottom risky networks, whilst being more divided in the middle section. It stands out that the perceived number one riskiest network, although clearly not objectively the riskiest, enjoys the highest unanimity amongst all participants.

Figure 6.4 Objective risk probability and unanimity for the full sample SMR ranking

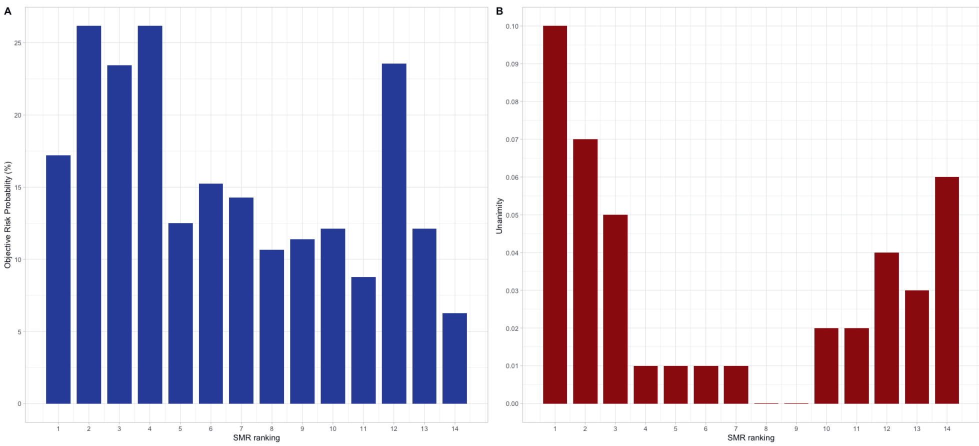


Table 6.6 shows the overall rankings of the 14 networks based on the observed subset mean ranks together with the subset unanimity of the rank frequencies. The corresponding

values of the observed subset mean ranks (per network and sample) appear in Table 6.9 in the Appendix. Also in this table, the last row shows the ranking of the networks based on the objective probability of infection. Similar to our model-based risk-utility rankings, we find noticeable differences between clusters. For instance, network ID 7 is ranked 12th by cluster 1, 14th by cluster 2, and 8th by cluster 3.

Like in Figure 6.4A, the degree of unanimity for each sample shows a distinct hyperbolic function in Table 6.6: both the top and bottom rankings display the highest level of unanimity. The middle sections of all samples display relatively low level of unanimity. This is explained by the fact that the most extreme rankings are achieved by the most extreme scores. Thus, agreement by the sample is needed to rank a network in the extreme spots of a ranking. Networks that end up in the middle section do so because of the lack of unanimity.

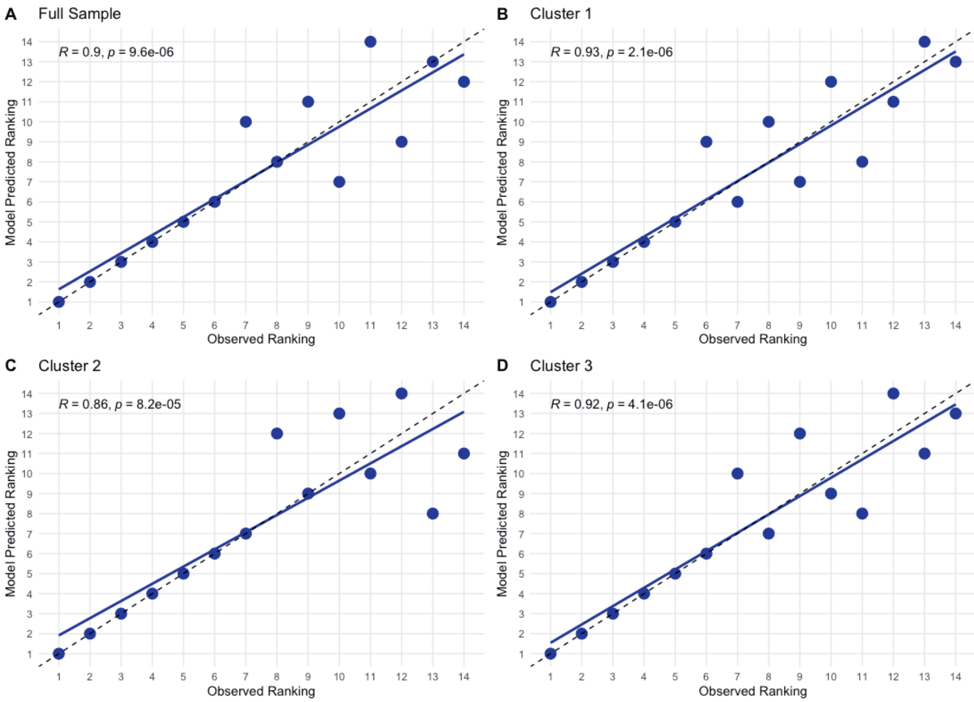
Table 6.6 Observed overall rankings and subset unanimity of the rank frequencies of the 14 networks, from most to least risky

Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Full Sample														
Network														
Unanimity	0.10	0.07	0.05	0.01	0.01	0.01	0.01	0.00	0.00	0.02	0.02	0.04	0.03	0.06
Cluster 1														
Network														
Unanimity	0.12	0.03	0.03	0.01	0.01	0.01	0.00	0.00	0.00	0.01	0.02	0.05	0.04	0.06
Cluster 2														
Network														
Unanimity	0.10	0.09	0.06	0.02	0.02	0.01	0.02	0.01	0.01	0.02	0.03	0.03	0.03	0.06
Cluster 3														
Network														
Unanimity	0.16	0.07	0.06	0.03	0.01	0.00	0.01	0.01	0.01	0.01	0.03	0.03	0.04	0.08
Objective Probability														
Network	 *	*							*	*				
ID	10	8	7	11	9	14	6	5	3	2	12	13	1	4

Note: For the row 'Objective Probability', * indicates that the two adjacent networks have the same risk probability and thus share their respective rank position.

We compare the observed overall rankings with the model-driven predicted rankings in Figure 6.5. Each value indicates the rank of a network in the model-driven predicted ranking (on the x-axis) and the rank that specific network has in the observed overall ranking (on the y-axis). For instance, Figure 6.5A shows the correlation between the two rankings for the full sample: the network that is predicted to rank 10th by our model, ranks 7th on the observed overall ranking. If our model were to perfectly predict the observed ranking, all values would fall on the dotted line (45° line). The straight line indicates the best fit between all values. The full model in figure 5A shows that the correlation between our predicted ranking and the observed ranking is relatively high (rho = .90, p<.001). Specifically, the first 6 out of the 14 rankings perfectly coincide with the model predicted ranking, based on the available characteristics.

Figure 6.5 Spearman's rank correlation between predicted and realized ranks



Note: The line indicates the linear fit between both ranks. The dotted line indicates perfect correlation.

The model's predictive power varies after splitting the sample into our identified clusters. Only for cluster 2 does the correlation decrease to .86 (p<.001), for clusters 1 and 3 the correlation between the predicted ranking and the observed ranking increases slightly (.93 and .92 respectively, p<.001). Figure 6.5B-D reveals that most deviation from perfect correlation resides in the low-rank section. The highest rankings, or highest (predicted) perceived risk, are most accurately predicted. Supported by our unanimity distribution, this leads to the conclusion that the model is most effective to predict unanimous preferences. When the participants vary in preference, our model loses predictive power.

When we compare the PML predictive ranking with the observed ranking, we notice first that the identified preference and weighing in the PML model is generally predictive of the observed behavior. A closer inspection of the correlations of both rankings with the objective risk probability in Table 6.7, suggests that if participants in cluster 3 would have systematically followed the PML model, they would have approximated the objective risk marginally better (yet, not perfectively). On the other hand, participants in cluster 2 are marginally better at approximating objective risk than the PML model based on their data predicted. Nevertheless, for the full sample, the PML model accurately predicts the observed ranking accuracy.

Table 6.7 Correlations of objective probability ranking with PML predicted ranking and Observed ranking

	Objective Probability	
	PML Prediction	Observation
Full Sample	0.74**	0.67**
Cluster 1	0.63*	0.59*
Cluster 2	0.56*	0.57*
Cluster 3	0.86***	0.85**

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

6.5 Discussion

This study explored how humans perceive risk spread in networks. Specifically, we asked participants to rate their connection to social networks with disease-infected contacts on the risk of infection. As to be expected, we show that the perceived risk preference does not perfectly coincide with the objective probability of risk to the decision-maker. Factors such as how close the closest infected contact is to the decider or the relation of infected to healthy network members, have comparable or stronger predictive values than the objective probability. The heterogeneity in the factor preference and weighing is shown when we identify three distinct subgroups using cluster analysis. The subgroup comparison suggests that demographic differences in education and gender may interact with the order and strength of perceived risk of these preferences, although our subgroups are not well enough identified to generalize these subgroup results based on this study. Furthermore, we show that we are able to relatively accurately predict the observed network preferences based on the preference estimates associated with the network characteristics.

Our results firmly consolidate that humans' processing of spread in social networks is not completely based on complex computations and depends on (more) easily observed characteristics of these networks. In line with bounded rationality, this does not imply that humans are unaware that the focus of risk assessment should be determining the objective risk or that this is physically impossible. We suggest that the often-complex mental calculation of objective risk dispersion in networks is substituted by a heuristics-driven approach. Simple observables such as the shortest distance to an infected node or the ratio of infected nodes relative to all nodes that are easily assessable become the basis of which individuals apply rules-of-thumb. This view is supported by the fact that the most important attributes are also relatively easily identified and accessible in real-life situations. Like all heuristic-based approaches, applying rules-of-thumb simplifies (and at times even ignores) complex aspects which can gravely influence the objective risk. For example, more a higher ratio of infections increase perceived risk, even in cases where objective risk is constant (e.g. compare networks with IDs 2 and 3. A higher ratio of infections is indeed perfectly correlated with increased risk in an unchanged network. However, even when we are able to keep the objective risk stable by varying other factors, a higher ratio of infections still increase the perceived risk. This example is depictive of our findings: heuristics have strong correlational power with objective risk in isolation, yet remain to resonate and dominate in complex networks even when other factors are also detrimental to the objective risk.

We conclude that the implications of our results are twofold. First, most practically, objective probabilities of social network spread are not easily computed. The relevance of this is directly applicable to the context in which our experiment is conducted. Throughout the COVID-19 crisis, the risk of infection was primarily communicated via a reproduction-metric (R). Our results suggest that only providing this metric without further training or guidance will result in a grossly inaccurate estimation of network dispersion. Policymakers trust blindly in the ability of individuals to apply this probability metric objectively, underestimating which other factors individuals actually rely on when connecting to social networks in order to make sense of this abstract probability. And this is not trivial: many countries like the Netherlands have largely left it to the citizens to decide with which social networks to connect (merely accompanied by recommendations). Providing examples of how such an R metric disperses infection chances amongst (complex) networks would have given the population a better handle to update their rules-of-thumb and apply them in real-life situations.

Second, our results are not limited to the contagion of infection. We have previously mentioned that the COVID-19 provided a clear and familiar context that enables us to translate a relatively abstract assessment into a relatable situation. This, however, does not mean that our results only apply to the risk of a pandemic-related infection in social networks. More broadly, our results identify general dispersion characteristics in social networks. Hence, the experienced closeness is relevant for personal and business contexts. Beyond the uncertainty of susceptibility and likelihood, the network structure of innovation, information, or any other social network will likewise influence the perceived willingness to connect to that network. For instance, pivotal research on the effect of (social) network structures on social capital hugely benefits from understanding the perceived spread in networks (Burt, 2017; Granovetter, 2018b, 2018a). Previously, the influence of network characteristics on perceived spread was obscured by other uncertainties related to network spread. Overall, identifying and understanding the underpinning of experienced social network strength and closeness in any (research) field opens the door to more tailored, effective, and generalizable policy, and economically relevant research design.

There are multiple limitations that warrant careful interpretation of our results. First, we have carefully constructed the networks by balancing a wide variety of network characteristics within a comparable range of size and risk probability. Effectively, multiple network variations within this range exist that were not included in our experiment. The conclusion of our paper itself, however, dictates that different variations or combinations of certain aspects could influence the order or strength of the estimates. For instance, our networks are relatively small: a maximum of 18 total nodes, 4 infected nodes, and 2 direct links limit our findings to compact networks. We cannot exclude the possibility that, for large networks, other heuristics apply or dominate. However, it is reasonable to assume that increasing the network, if anything, increases the dependency on heuristics instead of increasing accurate objective risk assessment. Ideally, the experiment would include more combinations of network characteristics, such as a (full) factorial set of combinations, grouped together in choice sets of size three. Due to practical limitations, we have focused on the most profound factors with a limited set of contrasting networks. Nevertheless, we underline throughout the paper that we interpret the main takeaway as being the clear indication that objective probability alone does not explain the risk preference in networks. The relative strength and order of the other influential factors are described, interpreted, and discussed, but not solidified as the definitive fixed effects in the population.

More research is needed to compare, estimate, and confirm the effect of different characteristics of networks on risk and network dispersion.

Second, infections are considered binary in the sense that a node is completely infected or not. In real life, asymptomatic people were still capable of spreading the virus undetected, and heavily coughing or sneezing infected people had been argued to spread the infection faster (El Hassan et al., 2022). We did not include asymptomatic dispersion of the virus: when a node got infected, it could infect the next one. Excluding asymptomatic dispersion variation therefore explicitly implies that there is no variation in the probability of infection. In our experiment, heavily infected and asymptomatic nodes spread the infection with the same 50% probability.

Third, individual differences have repeatedly proven to have a grave impact on risk assessment and risk aversion (e.g. Dohmen et al., 2011). We have partially attempted to control for individual differences by splitting our analysis into individual characteristics-based clusters. However, we cannot exclude that the true effect of individual characteristics on network risk assessment is more profound than we are able to identify within our cluster sample. Likewise, the context in which our risk assessment took place might have an effect on our results. We have framed the risk in networks as infection of COVID-19 within a social network. It is possible that the objective metrics we described (50% transmission rate) were either over- or underestimated based on individual experiences or beliefs about the spreading of the virus. We attempted to alleviate these concerns partially by excluding participants that reported a chance of infection different from ours. We further reran our analysis with controls that approximated the exposure using the deaths, hospitalizations, and positive tests due to COVID-19 at the postcode level. Although we found no significant effect of any of these proxies, we acknowledge that the possibility of differences at the individual level could have had a profound effect, remains.

Finally, our generalization assumes that the preference of heuristics is applied independent of context. The context of COVID-19, or disease more generally, could however constitute different preferred heuristics than innovation networks. Although we do not deem it likely that social network assessment in a different context depends less on heuristics, we cannot exclude the possibility that they weigh the importance of some characteristics differently. The specific ranking of important characteristics should be further confirmed in future research for different contexts.

6.6 Appendix

Table 6.8 Summary statistics of the total sample compared to the Dutch population

		Dutch population (2020)	Our sample (N=697)
Gender (Centraal Bureau voor de Statistiek, 2020)			
	Male	49.7%	56.0%
	Female	50.3%	44.0%
Age (van der Torre & Steenbekkers, 2020)			
	Mean in years	41.60[86]	42.09
Level of Education (van der Torre & Steenbekkers, 2020)			
	Low	20%	4.4%
	Medium	39%	33.1%
	High	41%	62.4%
Risk (Martín-Fernández et al., 2018)			
	General (0-10)	4,13[87]	5.26 (2.21)

[86] Measurement in 2014 included 26,5% of the age category 65+, an age group excluded from our sample.
[87] Most recent estimation in the Dutch population stems from 2018, amongst over 2,800 participants.

Table 6.9 MaxDiff PML model estimates: mean and standard deviations and significance of the attribute effects

Term	Category	Range	Mean estimate	SD	Subject SD	95% credible interval	p-value
Panel A: Main estimates							
Shortest path		(2, 5)	-0.665	0.088	0.354	[-0.83; -0.49]	0.006
Infected percentage		(0.091, 0.50)	1.518	0.068	0.213	[1.39; 1.66]	<0.000
Probability infection		(0.063, 0.262)	0.886	0.072	0.138	[0.76; 1.03]	<0.000
Number of direct links	1		-0.103	0.020	0.181	[-0.14; -0.06]	<0.000
	2		0.103	0.020	0.181	[-0.14; -0.06]	
Number of paths		(1, 16)	0.242	0.087	0.198	[0.06; 0.41]	0.882
Panel B: Cluster interaction estimates							
Shortest path							
	Cluster 1		0.196	0.094	0.304	[-0.03; 0.42]	
	Cluster 2		-1.578	0.109	1.121	[-1.85; -1.31]	<0.000
	Cluster 3		-0.613	0.101	0.276	[-0.90; -0.32]	
Infected percentage							
	Cluster 1		1.682	0.077	0.329	[1.53; 1.83]	
	Cluster 2		1.483	0.075	0.227	[1.28; 1.69]	<0.000
	Cluster 3		1.390	0.060	0.265	[1.17; 1.61]	
Probability infection							
	Cluster 1		0.775	0.072	0.130	[0.62; 0.93]	
	Cluster 2		0.563	0.075	0.226	[0.38; 0.75]	<0.000
	Cluster 3		1.321	0.069	0.221	[1.11; 1.53]	
Number of direct links							
	1 * Cluster 1		-0.103	0.020	0.181	[-0.14; -0.06]	
	2 * Cluster 2		-0.103	0.020	0.181	[-0.14; -0.06]	
	3 * Cluster 3		-0.103	0.020	0.181	[-0.14; -0.06]	>0.05
	1 * Cluster 1		-0.103	0.020	0.181	[-0.14; -0.06]	
	2 * Cluster 2		-0.103	0.020	0.181	[-0.14; -0.06]	
	3 * Cluster 3		-0.103	0.020	0.181	[-0.14; -0.06]	
Number of paths							
	Cluster 1		-0.075	0.114	0.234	[-0.29; 0.14]	
	Cluster 2		0.585	0.112	0.251	[0.33; 0.84]	<0.000
	Cluster 3		0.215	0.111	0.223	[-0.08; 0.51]	

Table 6.10 Ranking metrics per network per (sub)sample

Rankings			Ranking Metrics						
Observed Overall Rank	Utility Model Rank	Network ID	Objective Probability (Risk)	Mean Subset Rank	Subset Rank Unanimity	First	Second	Third	Marginal Risk Utility
Panel 1: Total sample									
1	1	9	17.19%	1.39	0.10	68.91%	23.36%	7.73%	1.55
2	2	10	26.17%	1.50	0.07	63.14%	23.54%	13.32%	1.11
3	3	11	23.44%	1.58	0.05	52.97%	36.36%	10.67%	1.04
4	4	8	26.17%	1.78	0.01	43.16%	35.63%	21.21%	-0.14
5	5	5	12.50%	1.85	0.01	41.47%	32.42%	26.11%	-0.17
6	6	14	15.23%	1.90	0.01	34.73%	40.69%	24.58%	-0.31
7	10	6	14.27%	2.00	0.01	27.77%	44.17%	28.07%	-1.52
8	8	13	10.64%	2.10	0.00	26.74%	36.52%	36.73%	-1.28
9	11	12	11.38%	2.12	0.00	28.71%	30.83%	40.46%	-1.79
10	7	3	12.11%	2.16	0.02	18.67%	46.95%	34.39%	-1.03
11	14	1	8.78%	2.26	0.02	22.93%	28.28%	48.78%	-2.00
12	9	7	23.57%	2.30	0.04	24.98%	19.76%	55.26%	-1.38
13	13	2	12.11%	2.35	0.03	12.75%	39.31%	47.94%	-1.83
14	12	4	6.25%	2.46	0.06	13.86%	26.47%	59.67%	-1.79
Panel 2: Cluster 1									
1	1	9	17.19%	1.34	0.12	73.22%	19.65%	7.13%	1.64
2	2	11	23.44%	1.66	0.03	47.03%	39.62%	13.35%	0.98
3	3	10	26.17%	1.68	0.03	50.11%	32.12%	17.77%	0.37
4	4	8	26.17%	1.84	0.01	40.20%	36.08%	23.73%	0.21
5	5	5	12.50%	1.86	0.01	42.19%	29.18%	28.62%	-0.07
6	9	6	14.27%	1.92	0.01	31.57%	44.42%	24.01%	-0.61
7	6	13	10.64%	1.97	0.00	33.96%	34.91%	31.13%	-0.17
8	10	12	11.38%	1.98	0.00	34.48%	32.76%	32.76%	-0.73
9	7	14	15.23%	2.02	0.00	28.92%	39.95%	31.13%	-0.35
10	12	1	8.78%	2.15	0.01	28.21%	28.80%	43.00%	-0.93
11	8	3	12.11%	2.17	0.02	18.48%	46.11%	35.41%	-0.46
12	11	7	23.57%	2.32	0.05	25.34%	17.08%	57.58%	-0.82
13	14	2	12.11%	2.36	0.04	12.22%	39.26%	48.52%	-1.34
14	13	4	6.25%	2.47	0.06	13.11%	26.72%	60.16%	-1.11
Panel 3: Cluster 2									
1	1	9	17.19%	1.39	0.10	68.97%	23.28%	7.76%	1.53
2	2	10	26.17%	1.47	0.09	68.04%	17.30%	14.66%	1.41
3	3	11	23.44%	1.52	0.06	57.78%	32.63%	9.58%	0.80
4	4	5	12.50%	1.75	0.02	46.31%	32.26%	21.43%	-0.10
5	5	14	15.23%	1.78	0.02	39.16%	44.01%	16.83%	-0.26
6	6	8	26.17%	1.79	0.01	40.99%	38.90%	20.10%	-1.02
7	7	3	12.11%	2.08	0.02	22.10%	47.77%	30.13%	-1.49
8	12	6	14.27%	2.09	0.01	23.71%	43.66%	32.63%	-2.47
9	9	13	10.64%	2.22	0.01	21.36%	35.25%	43.39%	-2.21
10	13	12	11.38%	2.29	0.02	22.19%	26.49%	51.32%	-2.74
11	10	2	12.11%	2.31	0.03	14.69%	39.86%	45.45%	-2.27
12	14	1	8.78%	2.32	0.03	20.33%	27.27%	52.39%	-2.86
13	8	4	6.25%	2.37	0.03	15.75%	31.74%	52.51%	-2.11
14	11	7	23.57%	2.45	0.06	17.30%	20.54%	62.16%	-2.42
Panel 4: Cluster 3									
1	1	10	26.17%	1.26	0.16	78.47%	17.01%	4.51%	1.73
2	2	9	17.19%	1.47	0.07	61.94%	29.41%	8.65%	1.62
3	3	11	23.44%	1.50	0.06	57.35%	35.29%	7.35%	1.52
4	4	8	26.17%	1.68	0.03	50.30%	31.10%	18.60%	0.16
5	5	14	15.23%	1.85	0.01	38.91%	37.66%	23.43%	-0.13
6	6	5	12.50%	1.93	0.00	34.84%	37.23%	27.93%	-0.13
7	10	6	14.27%	2.02	0.01	27.01%	44.39%	28.61%	-1.42
8	7	7	23.57%	2.09	0.01	33.89%	23.49%	42.62%	-0.92
9	12	12	11.38%	2.15	0.01	26.10%	32.35%	41.54%	-1.76
10	9	13	10.64%	2.18	0.01	20.66%	40.91%	38.43%	-1.32
11	8	3	12.11%	2.23	0.03	15.08%	47.24%	37.69%	-0.95
12	14	1	8.78%	2.36	0.03	17.80%	28.80%	53.40%	-2.07
13	11	2	12.11%	2.39	0.04	11.14%	38.71%	50.15%	-1.68
14	13	4	6.25%	2.53	0.08	13.04%	20.77%	66.18%	-1.94

Figure 6.6 Questionnaire part 1

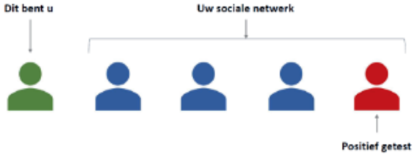
1. Experiment Introduction - Training Session

Page 1: *All questions will pertain to networks. These networks will be presented as follows:*

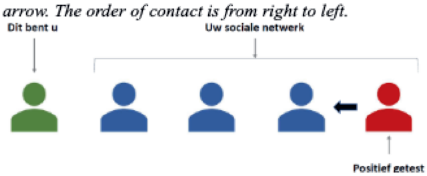
You are the green icon.

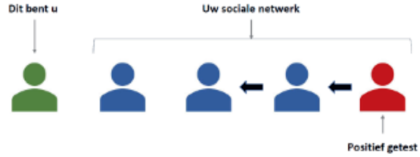
Page 2: *You will be presented with a schematic view of your social network. For example:*

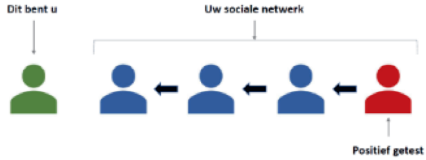
Page 3: *In every network, at least one person is infected. These people are red:*



Page 4: *You will see the order in which each person had contact with others. Contact is indicated by a black arrow. The order of contact is from right to left.*







Page 5. You will be shown three random networks. Please rank them: At the top, the network that you believe has the highest risk of infection for yourself (green). At the bottom, you place the network you believe has the lowest of all three networks. You can drag and drop accordingly. We are interested in your personal estimation.

The chance that an infection will pass to the next person in the network, is 50% each time.

Remember the people you (green) will connect to are indicated with a dotted line and a question mark. You will connect to all people if multiple dotted lines are indicated.

Summary:

- Rank each set of three network by setting the network that you believe has the highest risk of infection for yourself (green) at the top, and the network you believe has the lowest of all three networks at the bottom
- The chance that an infection will pass to the next person in the network, is 50% each time
- You (green) will connect to all people if multiple dotted lines are indicated.

For example:

In the network on the right, the risk to get infected for you (green) is 50%

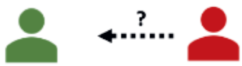


Figure 6.7 Questionnaire part 2

2. Level-of-understanding Assesment Questions

Question 1 (multiple choice):

If you connect directly to an infected person, the chance that you will get infected is:

- 25%
- 75%
- 50%
- 100%

Question 2 (multiple choice):

How are you asked to rank the three networks presented to you:

- The most risky networks at the top
- The least risky networks at the top
- The network with the most infected people at the top

Figure 6.8 Questionnaire part 3

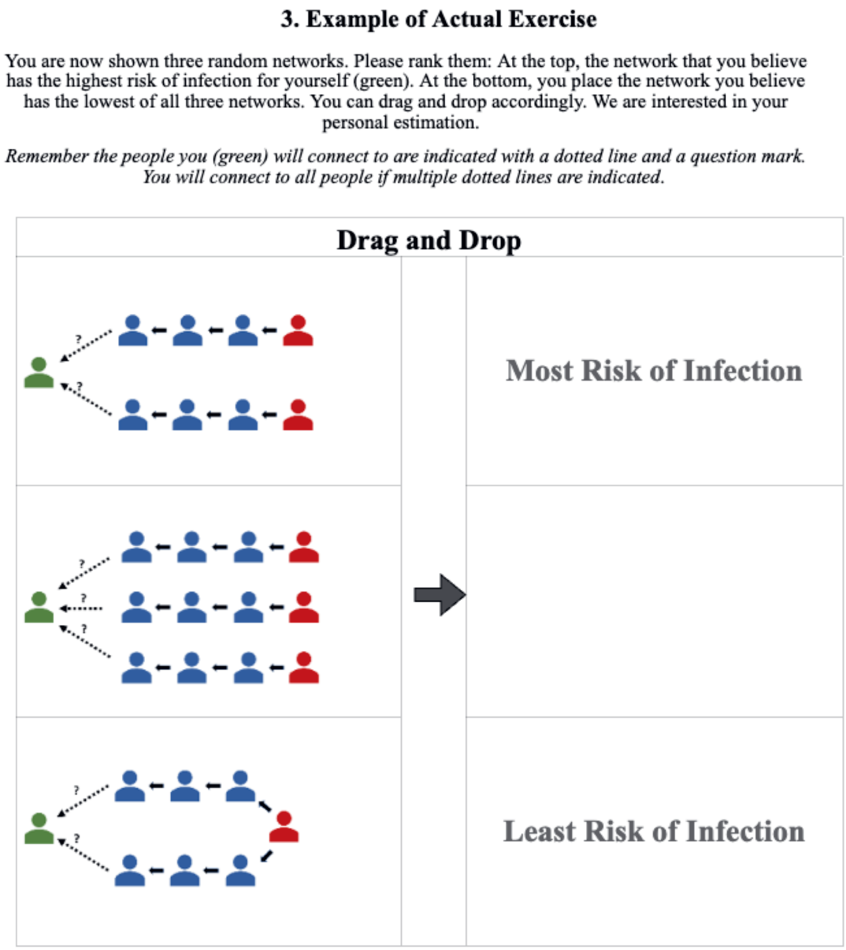
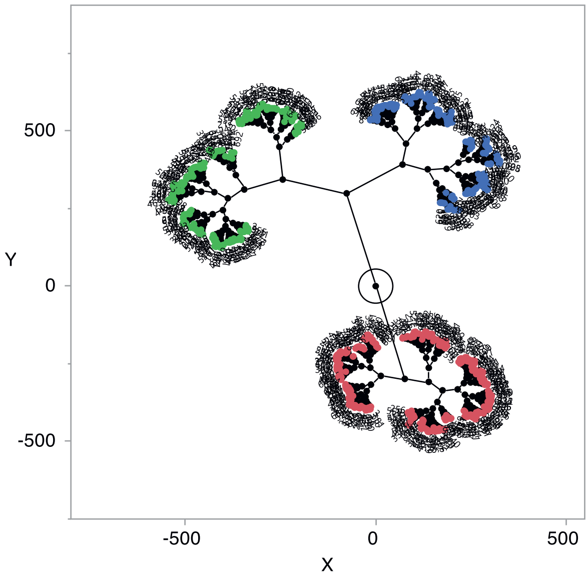


Figure 6.9 Constellation plot of cluster analysis



Note. Constellation plot of the 3 clusters of respondents showing the structure of the hierarchical clustering tree. Clusters 1, 2, and 3 represent respondents as red, green, and blue endpoints. The circle at the origin is the root of the tree, where the axes lay out the distance framework. Each cluster join is a new point and the lines represent membership in a cluster. The length of a line between cluster joins approximates the distance between the clusters that were joined.

Chapter 7

Summary of findings

In this dissertation, I set out to explore human behavior and thought processes with regard to (forced) changes in location decisions. I uncover the choice processes and choice consequences, as well as their evaluations. Moreover, I investigate the matter in which we evaluate our location choices. Throughout this dissertation, I use the pandemic as background, but my results are not limited to this context. It merely utilizes the necessity and novelty of the decision before us. Using that context, I find insights into how people deal with completely novel yet forceful policies, work from home, and how they (re)connect with risky social networks.

Chapter 2 examines how people decide to a decision to go out when the policy recommendation is to "avoid crowded places". Specifically, I look at the effect of context on the decision to visit a hypothetical recreational hotspot. I hypothesize that the absence of relevant up-to-date information about crowdedness will force individuals to make a decision based on unrelated information making it susceptible to biased reasoning. Using an experimental design, I show that people use expectations of others to influence their own decision: people go out when they expect to avoid crowded spots and when they expect others to go. The results suggest that in the former situation people act strategically, and in the latter social norms lead to escalation. "Use your common sense" is often the accompanying advice, but I show that more and better information concerning the context is essential to enable us to make optimal decisions for ourselves, and for society.

Chapter 3 has two distinct contributions to productivity research. First, I revise and validate a new Work Productivity and Stress Questionnaire (WPSQ) based on an older and less stable questionnaire. I show that the WPSQ factors outperform the older questionnaire on a large sample based on internal consistency and reliability on two measurement periods. I also show that single-item scale alternatives highly correlate with the subfactor productivity and stress (and irritability). Substituting the WPSQ factors with these single-scale alternatives should be done with caution and only when brevity demands it. Second, when I apply the WPSQ to working from home, I investigate trends over time in a volatile pandemic-ridden context. I argue that retrospective reports on work productivity and stress are strongly influenced by the current state. Specifically, the current state during the measurement structurally predicts the retrospective scores more than the targeted scores they aim to recollect. As a result, relying on retrospective scoring is subject to a recall bias and the current reference point should be taken into account.

Chapter 4 discusses to what extent satisfaction with home office hardware and the indoor environment influence productivity and burnout propensity during working from home (WFH). First, I find that self-reported productivity is higher at work compared to working from home. Second, participants prefer the indoor environment (e.g. temperature, air quality, lighting) at home over the environment in the work office, but prefer the work office hardware (e.g. screen, chair, Wi-Fi). Third, higher satisfaction with home environment factors significantly predicts increased productivity and decreased burnout propensity. Fourth, I connect real behavior with satisfaction scores and productivity. Increasing the amount of time spent in a ventilated room during working hours increases productivity and the willingness to continue WFH, whilst decreasing burnout propensity. This effect is fully mediated by satisfaction with the home office factors. Finally, I provide a strong case to emphasize actual measurement over self-reported satisfaction measurement. Ventilation influences related as well as unrelated factors' satisfaction scores. Consequently, satisfaction with unrelated aspects of the office, and thus WFH success, can be influenced (and improved) by seemingly unrelated actions such as increasing office ventilation. Taken together, this chapter shows that the physical climate in the home office influences the success of WFH.

Chapter 5 assesses the effect of indoor climate factors on human performance, focusing on the impact of indoor temperature on decision processes. Specifically, I expect heat to negatively influence higher cognitive rational processes, forcing people to rely more on intuitive shortcuts. In a laboratory setting, participants (N=257) were exposed to a controlled physical environment with either a hot temperature (28ºC) or a neutral temperature (22ºC), in which a battery of validated tests was conducted. I find that heat exposure did not lead to a difference in decision quality. I did find evidence for a strong gender difference in self-report, such that only men expect that high temperature leads to a significant decline in performance, which does in fact not materialize. These results cast doubt on the validity of self-report as a proxy for performance under different indoor climate conditions.

Chapter 6 uncovers how individuals perceive risk and spread through networks. I question whether individuals perceive some network structures as riskier than others. Since network spread can be relatively complex, I investigate whether humans' evaluation of the risk of COVID-19 spread through social networks is based on complex computations or whether it depends on (more) easily observed characteristics of these networks. I find that the perceived risk is not solely based on the objective probability of risk: easily assessable physical characteristics have stronger predictive values than the objective probability. This is in line with the theory of bounded rationality, where people are rationally motivated to optimize the problem, but limited mental capacity hampers or prevents them. The implications of this paper are not restricted to disease contagion. Physical characteristics could also partially predict the perceived spread in social networks such as reputation and fame in individual networks, or innovation or information in business networks.

Together, this dissertation contributes to a deeper understanding of how people make decisions in rapidly introduced, novel environments. I use experimental and data-driven evidence to unravel decisional processes, evaluations, and preferences during forced (re)location, often under uncertainty. I find evidence of both strategic and gut-feeling decision-making. In some contexts, these strategies flourish, and in others they underperform. Likewise, self-reported satisfaction or impact evaluations are often biased. The hidden common denominator throughout this dissertation is the novelty of the context. This novelty forces people to make a new assessment, with limited relevance of past experiences. The expectations or past experiences paint the evaluations and reflections, resulting in inaccuracy. Self-reported introspection and retrospection are at risk of being deceptive informants of internal decision processes and consequences in these novel decision problems.

Chapter 8

Societal Impact

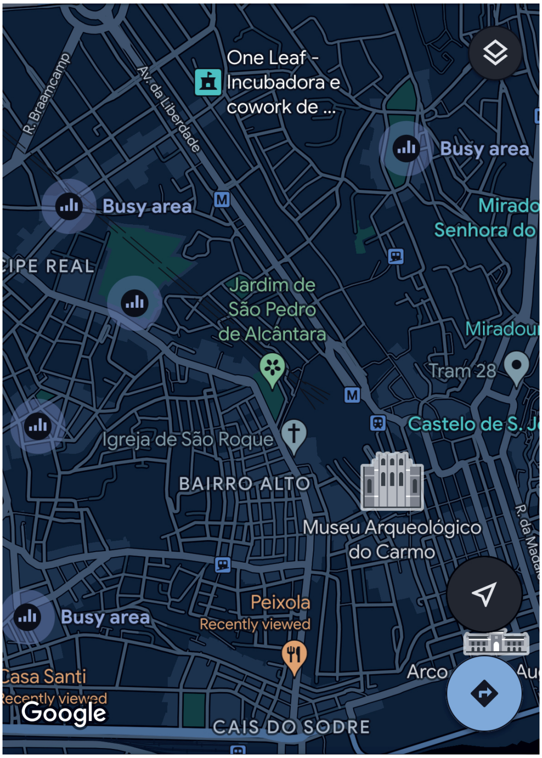
8.1 Impact on Social Policy

The recent period of pandemic-related crisis has seen the introduction of many novel policies. These policies aim to facilitate the needs of society, which often means finding a compromise between the often-conflicting public and personal preferences. Although the policy strategies to cater to these needs are carefully constructed, how these policies are received, interpreted, and executed by each individual is often unclear a priori. Two chapters from this dissertation evaluate abrupt policy implementations and help understand how people react to them.

Chapter 2 explains that the recommendation to avoid crowded areas is admirable, but leaves people to make their own prediction on what a crowded area is. The proxy people might use to estimate the crowdedness sometimes leads to a worse situation that is not in line with the needs of society or the individual. This is especially problematic because this dissertation shows that people generally have the intention to adhere to the recommendation. A future recommendation should be complemented with up-to-date information on the crowdedness level. Incidentally, a month after the publication of this chapter, Google released a new feature on Google Maps providing exactly this information: people are now able to see, live, how busy it is in popular areas (Moore, 2021).[88]

[88] After repeated contact with the Google Maps product manager, we were unable to publish a mutual press release on the societal impact of this feature due to privacy concerns voiced by users. Fortunately, I expect at least as many people to read this impact chapter.

Figure 8.1 Google Maps view of Lisbon indicating five 'Busy Areas'



However, providing up-to-date information does not automatically lead to accurate decision-making. For instance, the risk of infection was primarily communicated via an up-to-date reproduction metric (R; World Health Organization, 2020). From the results in chapter 6, I imply that only providing this metric without further training or guidance on how to use that information can result in an inaccurate estimation of network dispersion. Policymakers previously trusted blindly in the ability of individuals to apply this probability metric accurately. I unveil that individuals actually rely on other factors when they try to make sense of an abstract probability. Additionally, identifying and understanding the underpinning of experienced social network strength and closeness in any (research) field opens the door to more tailored, effective, and generalizable policy, and economically relevant research design.

Both chapters show that policies that largely leave the responsibility to the citizens to decide when to go out and whom to connect to, lack the understanding of the individual's decision process that follows. In other words, the intention of the Dutch government is justified, but the tools for optimal decision-making need to be improved. When future situations require similar recommendations to be put in place, up-to-date relevant information, as well as guidance, should be instated. This will enable individuals to act according to their preferences which, as this dissertation shows, is often also in accordance with societal interest.

8.2 Impact on Productivity, Health, and Well-being at home

Working from home is here to stay. Over 97% would like to continue to work from home, at least partially (Griffis, 2021). Employees are, on average, willing to take a 5% pay cut for 2-3 days of work from home (Aksoy et al., 2022) and in the future, 20% of all office work is predicted to be carried out from home (Etheridge et al., 2020). To ensure that this transition benefits workers, this dissertation set out to understand the effect of the physical home office on productivity, health, and well-being. A happy employee is a productive employee, and vice versa (e.g. Mamiseishvili & Rosser, 2011; McNeese-Smith, 1996; Miller & Monge, 1986). Chapter 4 concludes that, largely unbeknownst to homeworkers, the ventilation of the office improves satisfaction with the room, productivity, and willingness to continue working from home, as well as decreases the burnout propensity. Beyond the fact that ventilation is important, chapter 4 also underlines that it is likely that factors that are not directly in the focus of workers could also have a significant effect on satisfaction, health, and well-being. In contrast, chapter 5 shows that adverse conditions that are perceived to have a large influence on performance (in this case, heat) are overestimated.

Taken together, a tailored approach is needed to improve individual comfort, the quality of the home office, and its climate. Although this might not seem to be the most surprising conclusion, this thesis underlines that those improvements should not be solely focused on self-reports. Workers tend to consistently over- and underestimated to what extent factors actually influence their satisfaction and productivity. This statement contradicts a recent movement towards tailoring the indoor climate to (self-reported) comfort (Bluyssen, 2012, 2013). This dissertation questions the validity of convenient self-reported comfort as input for indoor climate interventions and emphasizes the need to contrast the self-report with objective measurements. In the end, a truly optimal physical climate is a key to the success and widely proposed bright future of working (from home).

8.3 Impact on Methodology

Multiple chapters of this dissertation examine the accuracy of self-reported data. Chapter 3 contributes to the research of productivity by attempting to validate (and as a result, reconstruct) an often-used questionnaire. This is, although arguably not sexy, an important endeavor. That the academic community is not eager to undergo this process is illustrated by the fact that the original development paper is, until now, cited 143 times, of which the majority of the papers passively administer the tool. Yet, the conclusion of that paper ends with a warning to not use the scale as a primary measure before additional validation has been conducted ("*Additional validation research on the HWQ is recommended before use as a primary measure in studies of worker productivity.*"; Shikiar et al., 2004, p. 226). Some of the citations compared the paper's factors to alternative tools, but none validate the survey. This dissertation produces a validated, new tool to measure productivity, stress, and other work-related factors.

Although unorthodox, the second methodological impact of this dissertation is to highlight the limitations of self-reports. The choice for self-reported measures is often dictated by a lack of objective alternatives, difficulties in collecting data, or general time constraints (Bloom & Van Reenen, 2007; Del Gatto et al., 2011; Färe et al., 1998; Gidwani & Dangayach, 2017; Singh et al., 2000; Skirbekk, 2004; Syverson, 2011).[89] For example, many of the working-from-home evaluations are based on self-reports (e.g. Aksoy et al., 2022; Barrero et al., 2021; Griffis, 2022). The future predictions based on these reports have a significant impact on companies' strategies to maintain their office real-estate (Gupta et al., 2022). But it appears that, especially in the economic domain, after selecting self-reported measures due to whatever constraints, the limitations are simply accepted and often passively mentioned as a footnote. To what extent these limitations drive the conclusions drawn from that data, instead of the actual effect, is hardly ever estimated.

Multiple chapters in this dissertation highlight how ignoring the limitations when effortlessly administrating a self-report tool translates into inaccurate conclusions. First, retrospectively reporting on relatively objective metrics such as productivity is subject to recall bias (Chapter 3).[90] Second, indoor environment factors influence satisfaction on unrelated

---

[89] The use of self-reported tools in this dissertation was often equally dictated by the lack of alternatives.

[90] "Relatively objective" refers to the extent that it should technically be objectively experienced or observable. This is in contrast to self-reporting an emotional state, which on its own could influence perception or judgment.

factors that appears to escape from conscious awareness. Consequent interventions aimed at improving these factors might show unsuccessful (Chapter 4). Third, how adverse indoor environment factors such as heat influence performance is overestimated (Chapter 5). These examples all point towards the value of combining self-report with objective data collection or measuring. This reasoning appears circular: self-reports are used when objective data is obscured, but objective data is needed to validate self-report accuracy. This dissertation suggests that future use of self-reports will benefit from a compromise. Acknowledging the limitations before collecting self-reported data should enable us to plan and collect additional, relevant, related data to check how robust and accurate the self-reports are. By checking how objective-related (or seemingly unrelated) measures correlated or predict self-reports, we will continue to develop our understanding of when self-reports are accurate, and when they are not.

References

Abraham, A., Hassanien, A. E., & Snášel, V. (2009). Computational Social Network Analysis: Trends, Tools and Research Advances. Springer London.

Adams-Prassl, A., Boneva, T., Golin, M., & Rauh, C. (2020). Inequality in the impact of the coronavirus shock: Evidence from real time surveys. Journal of Public Economics, 189, 104245. https://doi.org/10.1016/j.jpubeco.2020.104245

Adams-Prassl, A., Boneva, T., Golin, M., & Rauh, C. (2022). Work that can be done from home: Evidence on variation within and across occupations and industries. Labour Economics, 74, 102083. https://doi.org/10.1016/j.labeco.2021.102083

Adrjan, P., Ciminelli, G., Judes, A., Koelle, M., Schwellnus, C., & Sinclair, T. (2021). Will it stay or will it go? Analysing developments in telework during COVID-19 using online job postings data. 30. https://doi.org/10.1787/aed3816e-en

Aksoy, C. G., Barrero, J. M., Bloom, N., Davis, S. J., Dolls, M., & Zarate, P. (2022). Working from Home Around the World (Working Paper No. 30446). National Bureau of Economic Research. https://doi.org/10.3386/w30446

Al Horr, Y., Arif, M., Kaushik, A., Mazroei, A., Katafygiotou, M., & Elsarrag, E. (2016). Occupant productivity and office indoor environment quality: A review of the literature. Building and Environment, 105, 369–389. https://doi.org/10.1016/j.buildenv.2016.06.001

Allen, J. G., & Ibrahim, A. M. (2021). Indoor Air Changes and Potential Implications for SARS-CoV-2 Transmission. JAMA, 325(20), 2112–2113. https://doi.org/10.1001/jama.2021.5053

Allen, J. G., MacNaughton, P., Satish, U., Santanam, S., Vallarino, J., & Spengler, J. D. (2016). Associations of cognitive function scores with carbon dioxide, ventilation, and volatile organic compound exposures in office workers: A controlled exposure study of green and conventional office environments. Environmental Health Perspectives, 124(6), 805–812. https://doi.org/10.1289/ehp.1510037

Allen, T. D., Golden, T. D., & Shockley, K. M. (2015). How Effective Is Telecommuting? Assessing the Status of Our Scientific Findings. Psychological Science in the Public Interest, 16(2), 40–68. https://doi.org/10.1177/1529100615593273

Allen, V. L. (1965). Situational Factors In Conformity. Advances in Experimental Social Psychology, 2, 133–175. https://doi.org/10.1016/S0065-2601(08)60105-7

Amabile, T. M. (1988). A Model of Creativity and Innovation in Organizations. Research in Organizational Behavior, 10(1), 123. https://doi.org/Article

American Society of Heating, R. and A.-C. E. Inc. (2017). ASHRAE Handbook—Fundamentals (M. Owen, Ed.).

Anderson, C. A., Anderson, K. B., Dorr, N., DeNeve, K. M., & Flanagan, M. (2000). Temperature and Aggression. In Advances in Experimental Social Psychology (pp. 63–133). https://doi.org/10.1016/S0065-2601(00)80004-0

Arslan, G., Yıldırım, M., Tanhan, A., Buluş, M., & Allen, K.-A. (2021). Coronavirus stress, optimism-pessimism, psychological inflexibility, and psychological health: Psychometric properties of the Coronavirus Stress Measure. International Journal of Mental Health and Addiction, 19(6), 2423–2439.

Aumann, R. J., Maschler, M., & Stearns, R. E. (1995). Repeated games with incomplete information. MIT press.

Awada, M., Lucas, G., Becerik-Gerber, B., & Roll, S. (2021). Working from home during the COVID-19 pandemic: Impact on office worker productivity and work experience. Work, 69(4), 1171–1189. https://doi.org/10.3233/WOR-210301

Baars, B. J., & Franklin, S. (2003). How conscious experience and working memory interact. Trends in Cognitive Sciences, 7(4), 166–172. https://doi.org/10.1016/S1364-6613(03)00056-1

Barlas, Z., & Obhi, S. (2013). Freedom, choice, and the sense of agency. Frontiers in Human Neuroscience, 7. https://www.frontiersin.org/articles/10.3389/fnhum.2013.00514

Barone Gibbs, B., Kline, C., Huber, K., Paley, J., & Perera, S. (2021). COVID-19 shelter-at-home and work, lifestyle and well-being in desk workers. Occupational Medicine, 71(2), 86–94.

Barrero, J. M., Bloom, N., & Davis, S. J. (2021). Why Working from Home Will Stick (Working Paper No. 28731; Working Paper Series). National Bureau of Economic Research. https://doi.org/10.3386/w28731

Bartelsman, E. J., & Doms, M. (2000). Understanding Productivity: Lessons from Longitudinal Microdata. Journal of Economic Literature, 38(3), 569–594. https://doi.org/10.1257/jel.38.3.569

Baumeister, R., Bratslavsky, E., Muraven, M., & Tice, D. (1998). Ego Depletion: Is the active self a limited resource? Journal of Personality and Social Psychology, 74(3), 774–789. https://doi.org/10.1037/0022-3514.74.5.1252

Baumeister, R., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. Psychological Bulletin, 117(3), 497–529. https://doi.org/10.1037//0033-2909.117.3.497

Baumeister, R., Vohs, K. D., & Tice, D. M. (2007). The strength model of self-control. Current Directions in Psychological Science, 16(6), 351–355. https://doi.org/10.1111/j.1467-8721.2007.00534.x

Bazant, M. Z., & Bush, J. W. M. (2021). A guideline to limit indoor airborne transmission of COVID-19. Proceedings of the National Academy of Sciences, 118(17), e2018995118. https://doi.org/10.1073/pnas.2018995118

BBC News. (2020, May 21). Coronavirus: Alarm as crowds flock to European beaches. https://www.bbc.com/news/world-europe-52751228

Beaton, D., Bombardier, C., Escorpizo, R., Zhang, W., Lacaille, D., Boonen, A., Osborne, R. H., Anis, A. H., Strand, C. V., & Tugwell, P. S. (2009). Measuring Worker Productivity: Frameworks and Measures. The Journal of Rheumatology, 36(9), 2100–2109. https://doi.org/10.3899/jrheum.090366

Belanger, F. (1999). Workers' propensity to telecommute: An empirical study. Information & Management, 35, 139–153.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological), 57(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bialek, M., & Pennycook, G. (2017). The cognitive reflection test is robust to multiple exposures. Behavior Research Methods, 1–7. https://doi.org/10.3758/s13428-017-0963-x

Block, P., Hoffman, M., Raabe, I. J., Dowd, J. B., Rahal, C., Kashyap, R., & Mills, M. C. (2020). Social network-based distancing strategies to flatten the COVID-19 curve in a post-lockdown world. Nature Human Behaviour, 4(6), Article 6. https://doi.org/10.1038/s41562-020-0898-6

Bloom, N., Garicano, L., Sadun, R., & Van Reenen, J. (2014). The Distinct Effects of Information Technology and Communication Technology on Firm Organization. Management Science, 60(12), 2859–2885. https://doi.org/10.1287/mnsc.2014.2013

Bloom, N., Han, R., & Liang, J. (2022). How hybrid working from home works out. National Bureau of Economic Research.

Bloom, N., Liang, J., Roberts, J., & Ying, Z. J. (2015). Does working from home work? Evidence from a chinese experiment. Quarterly Journal of Economics, 130(1), 165–218. https://doi.org/10.1093/qje/qju032

Bloom, N., Schankerman, M., & Van Reenen, J. (2013). Identifying technology spillovers and product market rivalry. Econometrica, 81(4), 1347–1393.

Bloom, N., & Van Reenen, J. (2007). Measuring and Explaining Management Practices Across Firms and Countries*. The Quarterly Journal of Economics, 122(4), 1351–1408. https://doi.org/10.1162/qjec.2007.122.4.1351

Bluyssen, P. M. (2012). The Indoor Environment Handbook: How to make buildings healthy and comfortable. In Proceedings of the ICE - Engineering Sustainability (Vol. 165, Issue 3). Routledge. https://doi.org/10.1680/ensu.10.00054

Bluyssen, P. M. (2013). The Healthy Indoor Environment: How to Assess Occupants' Wellbeing in Buildings. Routledge. https://doi.org/10.4324/9781315887296

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. Frontiers in Public Health, 6, 149. https://doi.org/10.3389/fpubh.2018.00149

Bower, G. H. (1992). How might emotions affect learning. The Handbook of Emotion and Memory: Research and Theory, 3, 31.

Bowling, A. (2014). Research methods in health: Investigating health and health services. McGraw-hill education (UK).

Brañas-Garza, P., Kujal, P., & Lenkei, B. (2019). Cognitive reflection test: Whom, how, when. Journal of Behavioral and Experimental Economics, 82. https://doi.org/10.1016/j.socec.2019.101455

Braunack-Mayer, A., Tooher, R., Collins, J. E., Street, J. M., & Marshall, H. (2013). Understanding the school community's response to school closures during the H1N1 2009 influenza pandemic. BMC Public Health, 13(1), 344. https://doi.org/10.1186/1471-2458-13-344

Brennecke, J. (2020). Dissonant Ties in Intraorganizational Networks: Why Individuals Seek Problem-Solving Assistance from Difficult Colleagues. Academy of Management Journal, 63(3), 743–778. https://doi.org/10.5465/amj.2017.0399

Brennecke, J., & Rank, O. (2017). The firm's knowledge network and the transfer of advice among corporate inventors—A multilevel network study. Research Policy, 46(4), 768–783. https://doi.org/10.1016/j.respol.2017.02.002

Brooks, S. K., Webster, R. K., Smith, L. E., Woodland, L., Wessely, S., Greenberg, N., & Rubin, G. J. (2020). The psychological impact of quarantine and how to reduce it: Rapid review of the evidence. The Lancet, 395(10227), 912–920. https://doi.org/10.1016/S0140-6736(20)30460-8

Bruine de Bruin, W., & Bennett, D. (2020). Relationships Between Initial COVID-19 Risk Perceptions and Protective Health Behaviors: A National Survey. American Journal of Preventive Medicine, 59(2), 157–167. https://doi.org/10.1016/j.amepre.2020.05.001

Bruine de Bruin, W., Saw, H.-W., & Goldman, D. P. (2020). Political polarization in US residents' COVID-19 risk perceptions, policy preferences, and protective behaviors. Journal of Risk and Uncertainty, 61(2), 177–194. https://doi.org/10.1007/s11166-020-09336-3

Brunelle, E., & Fortin, J.-A. (2021). Distance Makes the Heart Grow Fonder: An Examination of Teleworkers' and Office Workers' Job Satisfaction Through the Lens of Self-Determination Theory. SAGE Open, 11(1), 2158244020985516. https://doi.org/10.1177/2158244020985516

Brutus, S., Aguinis, H., & Wassmer, U. (2013). Self-Reported Limitations and Future Directions in Scholarly Reports: Analysis and Recommendations. Journal of Management, 39(1), 48–75. https://doi.org/10.1177/0149206312455245

Bryant, H. E., Visser, N., & Love, E. J. (1989). Records, recall loss, and recall bias in pregnancy: A comparison of interview and medical records data of pregnant and postnatal women. American Journal of Public Health, 79(1), 78–80.

Brynjolfsson, E., Horton, J. J., Ozimek, A., Rock, D., Sharma, G., & TuYe, H.-Y. (2020). COVID-19 and Remote Work: An Early Look at US Data (Working Paper No. 27344; Working Paper Series). National Bureau of Economic Research. https://doi.org/10.3386/w27344

Buckee, C., Balsari, S., Chan, J., Crosas, M., Dominici, F., Gasser, U., Grad, Y. H., Grenfell, B., Halloran, M. E., Kraemer, M. U. G., Lipsitch, M., Metcalf, C. J. E., Meyers, L. A., Perkins, T. A., Santillana, M., Scarpino, S. V., Viboud, C., Wesolowski, A., & Schroeder, A. (2020). Aggregated mobility data could help fight COVID-19. Science (New York, N.Y.), 368(6487), 145–146. https://doi.org/10.1126/science.abb8021

Buckee, C., Noor, A., & Sattenspiel, L. (2021). Thinking clearly about social aspects of infectious disease transmission. Nature, 595(7866), Article 7866. https://doi.org/10.1038/s41586-021-03694-x

Bugental, D. B. (2000). Acquisition of the Algorithms of Social Life: A Domain-Based Approach. Psychological Bulletin, 126(2), 187–219. https://doi.org/10.1037/0033-2909.126.2.187

Bults, M., Beaujean, D. J., de Zwart, O., Kok, G., van Empelen, P., van Steenbergen, J. E., Richardus, J. H., & Voeten, H. A. (2011). Perceived risk, anxiety, and behavioural responses of the general public during the early phase of the Influenza A (H1N1) pandemic in the Netherlands: Results of three consecutive online surveys. BMC Public Health, 11(1), 2. https://doi.org/10.1186/1471-2458-11-2

Burgerperspectieven 2020 (publicatie No. 2; Burgerperspectieven). (2020). Sociaal en Cultureel Planbureau. https://www.scp.nl/publicaties/monitors/2020/06/29/burgerperspectieven-2020---2

Burt, R. S. (2017). Structural holes versus network closure as social capital. Social Capital, 31–56.

Byrne, N. M., Hills, A. P., Hunter, G. R., Weinsier, R. L., & Schutz, Y. (2005). Metabolic equivalent: One size does not fit all. Journal of Applied Physiology, 99(3), 1112–1119. https://doi.org/10.1152/japplphysiol.00023.2004

Byrnes, J. P., Miller, D. C., & Schafer, W. D. (1999). Gender Differences in Risk Taking: A Meta-Analysis. In Psychological Bulletin (Vol. 125, Issue 3, pp. 367–383).

Camerer, C. F., Ho, T. H., & Chong, J. K. (2004). A cognitive hierarchy model of games. Quarterly Journal of Economics, 119(3), 861–898. https://doi.org/10.1162/0033553041502225

Cao, M., & Wei, J. (2005). Stock market returns: A note on temperature anomaly. Journal of Banking and Finance, 29(6), 1559–1573. https://doi.org/10.1016/j.jbankfin.2004.06.028

Carter, E. C., Kofler, L. M., Forster, D. E., & Mccullough, M. E. (2015). A Series of Meta-Analytic Tests of the Depletion Effect: Self-Control Does Not Seem to Rely on a Limited Resource APA NLM. Association, 144(3), 0. https://doi.org/10.1037/xge0000083.supp

Cattell, R. B. (1966). The scree test for the number of factors. Multivariate Behavioral Research, 1(2), 245–276.

CBS. (2022). Alleenstaanden en gezinnen per gemeente (dashboard bevolking) [Webpagina]. Centraal Bureau voor Statistiek. https://www.cbs.nl/nl-nl/visualisaties/dashboard-bevolking/regionaal/alleenstaanden

Cedeño Laurent, J. G., Williams, A., Oulhote, Y., Zanobetti, A., Allen, J. G., & Spengler, J. D. (2018). Reduced cognitive function during a heat wave among residents of non-air-conditioned buildings: An observational study of young adults in the summer of 2016. PLOS Medicine, 15(7), e1002605. https://doi.org/10.1371/journal.pmed.1002605

Centola, D., & Macy, M. (2007). Complex Contagions and the Weakness of Long Ties. American Journal of Sociology, 113(3), 702–734. https://doi.org/10.1086/521848

Centraal Bureau voor de Statistiek. (2020). Mannen en vrouwen. Centraal Bureau voor de Statistiek. https://www.cbs.nl/nl-nl/visualisaties/dashboard-bevolking/mannen-en-vrouwen

Chang, T. Y., & Kajackaite, A. (2019). Battle for the thermostat: Gender and the effect of temperature on cognitive performance. PLoS ONE, 14(5). https://doi.org/10.1371/journal.pone.0216362

Cheema, A., & Patrick, V. M. (2012). Influence of Warm Versus Cool Temperatures on Consumer Choice: A Resource Depletion Account. Journal of Marketing Research, 49(6), 984–995. https://doi.org/10.1509/jmr.08.0205

Chinazzi, M., Davis, J. T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., Pastore y Piontti, A., Mu, K., Rossi, L., Sun, K., Viboud, C., Xiong, X., Yu, H., Halloran, M. E., Longini, I. M., & Vespignani, A. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science, 368(6489), 395–400. https://doi.org/10.1126/science.aba9757

Christakis, N. A., & Fowler, J. H. (2009). Connected: The surprising power of our social networks and how they shape our lives (1st ed). Little, Brown and Co.

Collins, C. (2020). Productivity in a pandemic. Science, 369(6504), 603–603.

Conway, M., & Ross, M. (1984). Getting what you want by revising what you had. Journal of Personality and Social Psychology, 47(4), 738.

Coovert, M. D., & Thompson, L. F. (2003). Technology and workplace health. In Handbook of occupational health psychology. (pp. 221–241). American Psychological Association. https://doi.org/10.1037/10474-011

Coughlin, S. S. (1990). Recall bias in epidemiologic studies. Journal of Clinical Epidemiology, 43(1), 87–91. https://doi.org/10.1016/0895-4356(90)90060-3

Craig, C., Overbeek, R. W., Condon, M. V., & Rinaldo, S. B. (2016). A relationship between temperature and aggression in NFL football penalties. Journal of Sport and Health Science, 5(2), 205–210. https://doi.org/10.1016/j.jshs.2015.01.001

Cross, T. J., Isautier, J. M., Stamatakis, E., Morris, S. J., Johnson, B. D., Wheatley-Guy, C., & Taylor, B. J. (2021). Self-reported physical activity before a COVID-19 'lockdown': Is it just a matter of opinion? BMJ Open Sport & Exercise Medicine, 7(2), e001088.

Cunningham, M. R., & Baumeister, R. F. (2016). How to Make Nothing Out of Something: Analyses of the Impact of Study Sampling and Statistical Interpretation in Misleading Meta-Analytic Conclusions. Frontiers in Psychology, 7, 1639. https://doi.org/10.3389/fpsyg.2016.01639

Curley, S. P., Yates, J. F., & Abrams, R. A. (1986). Psychological sources of ambiguity avoidance. Organizational Behavior and Human Decision Processes, 38(2), 230–256. https://doi.org/10.1016/0749-5978(86)90018-X

d'Andrea, V., Gallotti, R., Castaldo, N., & Domenico, M. D. (2022). Individual risk perception and empirical social structures shape the dynamics of infectious disease outbreaks. PLOS Computational Biology, 18(2), e1009760. https://doi.org/10.1371/journal.pcbi.1009760

Davis, K. G., Kotowski, S. E., Daniel, D., Gerding, T., Naylor, J., & Syck, M. (2020). The Home Office: Ergonomic Lessons From the "New Normal." Ergonomics in Design, 28(4), 4–10. https://doi.org/10.1177/1064804620937907

De Hond, M. (2020, October 4). Onderzoek 4 oktober over mondkapjes en angst om besmet te worden. https://maurice.nl/2020/10/04/onderzoek-4-oktober-over-mondkapjes-en-angst-om-besmet-te-worden/

De Neys, W., & Pennycook, G. (2019). Logic, Fast and Slow: Advances in Dual-Process Theorizing. Current Directions in Psychological Science, 28(5), 503–509. https://doi.org/10.1177/0963721419855658

De Vet, H. C., Terwee, C. B., Mokkink, L. B., & Knol, D. L. (2011). Measurement in medicine: A practical guide. Cambridge university press.

Del Gatto, M., Di Liberto, A., & Petraglia, C. (2011). Measuring productivity. Journal of Economic Surveys, 25(5), 952–1008.

Demirtaş-Madran, H. A. (2021). Accepting Restrictions and Compliance With Recommended Preventive Behaviors for COVID-19: A Discussion Based on the Key Approaches and Current Research on Fear Appeals. Frontiers in Psychology, 12, 558437. https://doi.org/10.3389/fpsyg.2021.558437

Denson, T. F., DeWall, C. N., & Finkel, E. J. (2012). Self-control and aggression. Current Directions in Psychological Science, 21(1), 20–25. https://doi.org/10.1177/0963721411429451

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. Journal of Abnormal and Social Psychology, 51(3), 629–636. https://doi.org/10.1037/h0046408

Devine, D., Gaskell, J., Jennings, W., & Stoker, G. (2021). Trust and the Coronavirus Pandemic: What are the Consequences of and for Trust? An Early Review of the Literature. Political Studies Review, 19(2), 274–285. https://doi.org/10.1177/1478929920948684

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. Journal of the European Economic Association, 9(3), 522–550. https://doi.org/10.1111/j.1542-4774.2011.01015.x

Drews, C. D., & Greeland, S. (1990). The Impact of Differential Recall on the Results of Case-Control Studies. International Journal of Epidemiology, 19(4), 1107–1112. https://doi.org/10.1093/ije/19.4.1107

Dryhurst, S., Schneider, C. R., Kerr, J., Freeman, A. L. J., Recchia, G., van der Bles, A. M., Spiegelhalter, D., & van der Linden, S. (2020). Risk perceptions of COVID-19 around the world. Journal of Risk Research, 23(7–8), 994–1006. https://doi.org/10.1080/13669877.2020.1758193

Easley, D., & Kleinberg, J. (2010). Networks, crowds, and markets: Reasoning about a highly connected world. Cambridge university press.

Eichholtz, P., Holtermans, R., & Kok, N. (2019). Environmental Performance of Commercial Real Estate: New Insights into Energy Efficiency Improvements. Journal of Portfolio Management, 45(7), 113–129. https://doi.org/10.3905/jpm.2019.1.099

El Hassan, M., Assoum, H., Bukharin, N., Al Otaibi, H., Mofijur, M., & Sakout, A. (2022). A review on the transmission of COVID-19 based on cough/sneeze/breath flows. European Physical Journal Plus, 137(1), 1. https://doi.org/10.1140/epjp/s13360-021-02162-9

Engelbert, M., & Carruthers, P. (2010). Introspection. Wiley Interdisciplinary Reviews: Cognitive Science, 1(2), 245–253. https://doi.org/10.1002/wcs.4

Engle, R. W., & Kane, M. J. (2003). Executive Attention, Working Memory Capacity, and a Two-Factor Theory of Cognitive Control. Psychology of Learning and Motivation - Advances in Research and Theory, 44, 145–199. https://doi.org/10.1016/S0079-7421(03)44005-X

Escorpizo, R., Bombardier, C., Boonen, A., Hazes, J. M., Lacaille, D., Strand, V., & Beaton, D. (2007). Worker productivity outcome measures in arthritis. The Journal of Rheumatology, 34(6), 1372–1380.

Etheridge, B., Tang, L., & Wang, Y. (2020). Worker productivity during lockdown and working from home: Evidence from self-reports. Covid Economics, 52, 118–151.

Evans, J. S. (2007). On the resolution of conflict in dualprocess theories of reasoning. Thinking and Reasoning, 13(4), 321–329.

Falk, A., Dohmen, T., & Huffman, D. (2016). The preference survey module: A validated instrument for measuring risk, time, and social preferences. IZA Discussion Paper, No. 9674(9674).

Fang, H., Wang, L., & Yang, Y. (2020). Human mobility restrictions and the spread of the Novel Coronavirus (2019-nCoV) in China. Journal of Public Economics, 191, 104272. https://doi.org/10.1016/j.jpubeco.2020.104272

Fang, L., Wyon, D. P., Clausen, G., & Fanger, P. O. (2004). Impact of indoor air temperature and humidity in an office on perceived air quality, SBS symptoms and performance. Indoor Air, 14(s7), 74–81. https://doi.org/10.1111/j.1600-0668.2004.00276.x

Färe, R., Grosskopf, S., & Roos, P. (1998). Malmquist productivity indexes: A survey of theory and practice. In Index numbers: Essays in honour of Sten Malmquist (pp. 127–190). Springer.

Farooq, R., & Sultana, A. (2021). The potential impact of the COVID-19 pandemic on work from home and employee productivity. Measuring Business Excellence.

Felstead, A., & Henseke, G. (2017). Assessing the growth of remote working and its consequences for effort, well-being and work-life balance. New Technology, Work and Employment, 32(3), 195–212. https://doi.org/10.1111/ntwe.12097

Ferketich, S. (1991). Focus on psychometrics. Aspects of item analysis. Research in Nursing & Health, 14(2), 165–168.

Festinger, L. (1957). A theory of cognitive dissonance. In A Theory of Cognitive Dissonance (Vol. 2). Stanford University Press. https://doi.org/10.4324/9781912282432

Frederick, S. (2005). Cognitive Reflection and Decision Making. Journal of Economic Perspectives, 19(4), 25–42. https://doi.org/10.1257/089533005775196732

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. Social Networks, 1(3), 215–239.

Frey, R., Pedroni, A., Mata, R., Rieskamp, J., & Hertwig, R. (2017). Risk preference shares the psychometric structure of major psychological traits. Science Advances, 3(10). https://doi.org/10.1126/sciadv.1701381

Frontczak, M., & Wargocki, P. (2011). Literature survey on how different factors influence human comfort in indoor environments. Building and Environment, 46(4), 922–937. https://doi.org/10.1016/j.buildenv.2010.10.021

Funk, S., Bansal, S., Bauch, C. T., Eames, K. T. D., Edmunds, W. J., Galvani, A. P., & Klepac, P. (2015). Nine challenges in incorporating the dynamics of behaviour in infectious diseases models. Epidemics, 10, 21–25. https://doi.org/10.1016/j.epidem.2014.09.005

Gajendran, R. S., & Harrison, D. A. (2007). The good, the bad, and the unknown about telecommuting: Meta-analysis of psychological mediators and individual consequences. Journal of Applied Psychology, 92(6), 1524–1541. https://doi.org/10.1037/0021-9010.92.6.1524

Gajendran, R. S., Harrison, D. A., & Delaney-Klinger, K. (2015). Are Telecommuters Remotely Good Citizens? Unpacking Telecommuting's Effects on Performance. Personnel Psychology, 68(2), 353–393. https://doi.org/10.1111/peps.12082

Galanti, T., Guidetti, G., Mazzei, E., Zappalà, S., & Toscano, F. (2021). Work from home during the COVID-19 outbreak: The impact on employees' remote work productivity, engagement, and stress. Journal of Occupational and Environmental Medicine, 63(7), e426.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. Psychological Bulletin, 132(5), 692–731. https://doi.org/10.1037/0033-2909.132.5.692

Gibbs, M., Mengel, F., & Siemroth, C. (2021). Work from Home & Productivity: Evidence from Personnel & Analytics Data on IT Professionals (Working Paper No. 14336). IZA Discussion Papers.

Gidwani, B. D., & Dangayach, G. (2017). Productivity measurement and improvement-an overview. International Journal of Productivity and Quality Management, 20(3), 316–343.

Gigerenzer, G., & Gaissmaier, W. (2015). Decision making: Nonrational theories. In International encyclopedia of the social & behavioral sciences (pp. 911–916). Elsevier.

Gigerenzer, G., Hertwig, R., & Pachur, T. (Eds.). (2011). Heuristics: The Foundations of Adaptive Behavior. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199744282.001.0001

Gigerenzer, G., & Selten, R. (2002). Bounded rationality: The adaptive toolbox. MIT press.

Glass, R. J., Glass, L. M., Beyeler, W. E., & Min, H. J. (2006). Targeted social distancing designs for pandemic influenza. Emerging Infectious Diseases, 12(11), 1671.

Golden, T. D. (2006). Avoiding depletion in virtual work: Telework and the intervening impact of work exhaustion on commitment and turnover intentions. Journal of Vocational Behavior, 69(1), 176–187. https://doi.org/10.1016/j.jvb.2006.02.003

Golden, T. D., & Veiga, J. F. (2005). The Impact of Extent of Telecommuting on Job Satisfaction: Resolving Inconsistent Findings. Journal of Management, 31(2), 301–318. https://doi.org/10.1177/0149206304271768

Golden, T. D., Veiga, J. F., & Dino, R. N. (2008). The impact of professional isolation on teleworker job performance and turnover intentions: Does time spent teleworking, interacting face-to-face, or having access to communication-enhancing technology matter? Journal of Applied Psychology, 93(6), 1412–1421. https://doi.org/10.1037/a0012722

Granovetter, M. (2018a). Economic action and social structure: The problem of embeddedness. In The sociology of economic life (pp. 22–45). Routledge.

Granovetter, M. (2018b). The impact of social structure on economic outcomes. In The Sociology of Economic Life (pp. 46–61). Routledge.

Grassly, N. C., & Fraser, C. (2008). Mathematical models of infectious disease transmission. Nature Reviews Microbiology, 6(6), Article 6. https://doi.org/10.1038/nrmicro1845

Grether, W. F. (1973). Human performance at elevated environmental temperatures. Aerospace Medicine, 44(7), 747–755.

Griffis, H. (2021). State of Remote Work 2021 (State of Remote Work). Buffer. https://buffer.com/2021-state-of-remote-work

Griffis, H. (2022). State Of Remote Work 2022 (State of Remote Work). Buffer. https://buffer.com/state-of-remote-work/2022

Grove, S. K., Burns, N., & Gray, J. (2012). The practice of nursing research: Appraisal, synthesis, and generation of evidence. Elsevier Health Sciences.

Gupta, A., Mittal, V., Peeters, J., & Van Nieuwerburgh, S. (2022). Flattening the curve: Pandemic-Induced revaluation of urban real estate. Journal of Financial Economics, 146(2), 594–636. https://doi.org/10.1016/j.jfineco.2021.10.008

Hafenbrädl, S., Waeger, D., Marewski, J. N., & Gigerenzer, G. (2016). Applied decision making with fast-and-frugal heuristics. Journal of Applied Research in Memory and Cognition, 5(2), 215–231.

Hagen, L., Keller, T., Neely, S., DePaula, N., & Robert-Cooperman, C. (2018). Crisis Communications in the Age of Social Media: A Network Analysis of Zika-Related Tweets. Social Science Computer Review, 36(5), 523–541. https://doi.org/10.1177/0894439317721985

Hagger, M. S., Chatzisarantis, N. L. D., Alberts, H., Anggono, C. O., Batailler, C., Birt, A. R., Brand, R., Brandt, M. J., Brewer, G., Bruyneel, S., Calvillo, D. P., Campbell, W. K., Cannon, P. R., Carlucci, M., Carruth, N. P., Cheung, T., Crowell, A., De Ridder, D. T. D., Dewitte, S., … Zwienenberg, M. (2016). A Multilab Preregistered Replication of the Ego-Depletion Effect. Perspectives on Psychological Science, 11(4), 546–573. https://doi.org/10.1177/1745691616652873

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. Psychological Review, 108(4), 814–834. https://doi.org/10.1037/0033-295X.108.4.814

Halkos, G., & Bousinakis, D. (2010). The effect of stress and satisfaction on productivity. International Journal of Productivity and Performance Management.

Hallman, D. M., Januario, L. B., Mathiassen, S. E., Heiden, M., Svensson, S., & Bergström, G. (2021). Working from home during the COVID-19 outbreak in Sweden: Effects on 24-h time-use in office workers. BMC Public Health, 21(1), 528. https://doi.org/10.1186/s12889-021-10582-6

Halpern, M. T., Shikiar, R., Rentz, A. M., & Khan, Z. M. (2001). Impact of smoking status on workplace absenteeism and productivity. Tobacco Control, 10(3), 233. https://doi.org/10.1136/tc.10.3.233

Hancock, P. A., & Vasmatzidis, I. (1998). Human occupational and performance limits under stress: The thermal environment as a prototypical example. Ergonomics, 41(8), 1169–1191. https://doi.org/10.1080/001401398186469

Hancock, P. A., & Vasmatzidis, I. (2003). Effects of heat stress on cognitive performance: The current state of knowledge. International Journal of Hyperthermia, 19, 355–372. https://doi.org/10.1080/0265673021000054630

Haney, C., Banks, C., & Zimbardo, P. (1973). Interpersonal dynamics in a simulated prison. International Journal of Criminology & Penology, 1(1), 69–97.

Hansen, S., Lambert, P. J., Bloom, N., Davis, S. J., Sadun, R., & Taska, B. (2022). Remote Work across Jobs, Companies, and Countries. working paper, July.

Hardin, G. (1968). The tragedy of the commons. Science, 162(3859), 1243–1248. https://doi.org/10.1126/science.162.3859.1243

He, S., Chen, S., Kong, L., & Liu, W. (2021). Analysis of Risk Perceptions and Related Factors Concerning COVID-19 Epidemic in Chongqing, China. Journal of Community Health, 46(2), 278–285. https://doi.org/10.1007/s10900-020-00870-4

Healy, G. N., Eakin, E. G., Owen, N., LaMontagne, A. D., Moodie, M., Winkler, E. A., Fjeldsoe, B. S., Wiesner, G., Willenberg, L., & Dunstan, D. W. (2016). A cluster randomized controlled trial to reduce office workers' sitting time: Impact on activity outcomes. Medicine & Science in Sports & Exercise, 48(9), 1787–1797.

Hermalin, B. E., & Weisbach, M. S. (1991). The Effects of Board Composition and Direct Incentives on Firm Performance. Financial Management, 20(4), 101. https://doi.org/10.2307/3665716

Ho, T. H., & Su, X. (2013). A dynamic level-K model in sequential games. Management Science, 59(2), 452–469. https://doi.org/10.1287/mnsc.1120.1645

Hockey, R. (2013). The psychology of fatigue: Work, effort and control. In The Psychology of Fatigue: Work, Effort and Control. Cambridge Univ. Press. https://doi.org/10.1017/CBO9781139015394

Holmberg, L., Ohlander, E., Byers, T., Zack, M., Wolk, A., Bruce, Å., Bergstrōm, R., Bergkvist, L., & Adami, H.-O. (1996). A search for recall bias in a case-control study of diet and breast cancer. International Journal of Epidemiology, 25(2), 235–244.

Holt, C. A., & Laury, S. K. (2002). Risk Aversion and Incentive Effects. American Economic Review, 92(5), 1644–1655. https://doi.org/10.1257/000282802762024700

Holt-Lunstad, J., & Smith, T. B. (2012). Social relationships and mortality. Social and Personality Psychology Compass, 6(1), 41–53. https://doi.org/10.1111/j.1751-9004.2011.00406.x

Hovinga, C. A., Asato, M. R., Manjunath, R., Wheless, J. W., Phelps, S. J., Sheth, R. D., Pina-Garza, J. E., Zingaro, W. M., & Haskins, L. S. (2008). Association of non-adherence to antiepileptic drugs and seizures, quality of life, and productivity: Survey of patients with epilepsy and physicians. Epilepsy & Behavior, 13(2), 316–322.

Huizenga, C., Abbaszadeh, S., Zagreus, L., & Arens, E. (2006). Air quality and thermal comfort in office buildings: Results of a large indoor environmental quality survey. Proceedings of Healthy Buildings, III, 393–397. https://doi.org/10.12659/PJR.894050

Huremović, D. (Ed.). (2019). Psychiatry of pandemics: A mental health response to infection outbreak. Springer. https://doi.org/10.1007/978-3-030-15346-5

Inglehart, R., Foa, R., Peterson, C., & Welzel, C. (2008). Development, Freedom, and Rising Happiness: A Global Perspective (1981–2007). Perspectives on Psychological Science, 3(4), 264–285. https://doi.org/10.1111/j.1745-6924.2008.00078.x

Ingram, K. M., Mickes, L., & Wixted, J. T. (2012). Recollection can be weak and familiarity can be strong. Journal of Experimental Psychology: Learning, Memory, and Cognition, 38(2), 325–339. https://doi.org/10.1037/a0025483

Inzlicht, M., Werner, K. M., Briskin, J. L., & Roberts, B. W. (2021). Integrating Models of Self-Regulation. Annual Review of Psychology, 72, 319–345. https://doi.org/10.1146/annurev-psych-061020-105721

Ipsen, C., van Veldhoven, M., Kirchner, K., & Hansen, J. P. (2021). Six Key Advantages and Disadvantages of Working from Home in Europe during COVID-19. International Journal of Environmental Research and Public Health, 18(4). https://doi.org/10.3390/ijerph18041826

Ishii, K., Shibata, A., & Oka, K. (2018). Work Engagement, Productivity, and Self-Reported Work-Related Sedentary Behavior Among Japanese Adults: A Cross-Sectional Study. Journal of Occupational and Environmental Medicine, 60(4), e173. https://doi.org/10.1097/JOM.0000000000001270

Jeong, H., Yim, H. W., Song, Y.-J., Ki, M., Min, J.-A., Cho, J., & Chae, J.-H. (2016). Mental health status of people isolated due to Middle East Respiratory Syndrome. Epidemiology and Health, 38, e2016048. https://doi.org/10.4178/epih.e2016048

Johns, G. (2010). Presenteeism in the workplace: A review and research agenda. Journal of Organizational Behavior, 31(4), 519–542. https://doi.org/10.1002/job.630

Johnson, J. (2001). Evaluation of learning according to objectives tool. Springer Publishing Company, New York.

Joordens, S., & Hockley, W. E. (2000). Recollection and familiarity through the looking glass: When old does not mirror new. Journal of Experimental Psychology: Learning, Memory, and Cognition, 26(6), 1534–1555. https://doi.org/10.1037/0278-7393.26.6.1534

Kahn, M. E. (2009). Urban growth and climate change. Annual Review of Resource Economics, 1(1), 333–350.

Kahneman, D. (1973). Attention and Effort. Englewood Cliffs.

Kahneman, D., & Frederick, S. (2005). A model of heuristic judgment. In K. Holyoak & B. Morrison (Eds.), The Cambridge Handbook of Thinking and Reasoning (pp. 267--293). Cambridge University Press.

Kahneman, D., Slovic, S. P., Slovic, P., & Tversky, A. (1982). Judgment under uncertainty: Heuristics and biases. Cambridge university press.

Karasneh, R., Al-Azzam, S., Muflih, S., Soudah, O., Hawamdeh, S., & Khader, Y. (2021). Media's effect on shaping knowledge, awareness risk perceptions and communication practices of pandemic COVID-19 among pharmacists. Research in Social & Administrative Pharmacy, 17(1), 1897–1902. https://doi.org/10.1016/j.sapharm.2020.04.027

Keizer, K., Lindenberg, S., & Steg, L. (2008). The spreading of disorder. Science, 322(5908), 1681–1685. https://doi.org/10.1126/science.1161405

Kerr, N. L., Rumble, A. C., Park, E. S., Ouwerkerk, J. W., Parks, C. D., Gallucci, M., & van Lange, P. A. M. (2009). "How many bad apples does it take to spoil the whole barrel?": Social exclusion and toleration for bad apples. Journal of Experimental Social Psychology, 45(4), 603–613. https://doi.org/10.1016/j.jesp.2009.02.017

Kim, J., & de Dear, R. (2013). Workspace satisfaction: The privacy-communication trade-off in open-plan offices. Journal of Environmental Psychology, 36, 18–26. https://doi.org/10.1016/j.jenvp.2013.06.007

Kingma, B., & Van Marken Lichtenbelt, W. (2015). Energy consumption in buildings and female thermal demand. Nature Climate Change, 5(12), 1054–1056. https://doi.org/10.1038/nclimate2741

Klein, C. T. F., & Helweg-Larsen, M. (2002). Perceived Control and the Optimistic Bias: A Meta-Analytic Review. Psychology & Health, 17(4), 437–446. https://doi.org/10.1080/0887044022000004920

Kline, P. (1993). Handbook of psychological testing. Routledge.

Korolevich, Sara. (2021). The State Of Remote Work In 2021: A Survey Of The American Workforce. Goodhire. https://www.goodhire.com/resources/articles/state-of-remote-work-survey/

Krekel, C., Ward, G., & De Neve, J.-E. (2019). Employee well-being, productivity, and firm performance: Evidence and case studies. Global Happiness and Wellbeing.

Künn, S., Palacios, J., & Pestel, N. (2019). The Impact of Indoor Climate on Human Cognition: Evidence from Chess Tournaments. IZA Discussion Paper, 12632, 41.

Lamberson, P. J. (2016). Diffusion in networks. In Y. Bramoullé, A. Galeotti, & B. W. Rogers (Eds.), The Oxford handbook of the economics of networks. Oxford University Press.

Lan, L., & Lian, Z. (2009). Use of neurobehavioral tests to evaluate the effects of indoor environment quality on productivity. Building and Environment, 44(11), 2208–2217. https://doi.org/10.1016/j.buildenv.2009.02.001

Lan, L., Lian, Z., Pan, L., & Ye, Q. (2009). Neurobehavioral approach for evaluation of office workers' productivity: The effects of room temperature. Building and Environment, 44(8), 1578–1588. https://doi.org/10.1016/j.buildenv.2008.10.004

Lee, F., Teich, J. M., Spurr, C. D., & Bates, D. W. (1996). Implementation of Physician Order Entry: User Satisfaction and Self-Reported usage Patterns. Journal of the American Medical Informatics Association, 3(1), 42–55. https://doi.org/10.1136/jamia.1996.96342648

Leith, K. P., & Baumeister, R. F. (1996). Why do bad moods increase self-defeating behavior? Emotion, risk taking, and self-regulation. Journal of Personality and Social Psychology, 71(6), 1250–1267. https://doi.org/10.1037/0022-3514.71.6.1250

Liebrand, W., Messick, D., & Wilke, H. (Eds.). (1992). Goals, Expectations and Behavior in a Social Dilemma Situation. In Social Dilemmas. Garland Science.

Lin, H., Saunders, B., Friese, M., Evans, N. J., & Inzlicht, M. (2020). Strong Effort Manipulations Reduce Response Caution: A Preregistered Reinvention of the Ego-Depletion Paradigm. Psychological Science, 31(5), 531–547. https://doi.org/10.1177/0956797620904990

Liu, M., Zhang, H., & Huang, H. (2020). Media exposure to COVID-19 information, risk perception, social and geographical proximity, and self-rated anxiety in China. BMC Public Health, 20(1), 1649. https://doi.org/10.1186/s12889-020-09761-8

Lofland, J. H., Pizzi, L., & Frick, K. D. (2004). A Review of Health-Related Workplace Productivity Loss Instruments. PharmacoEconomics, 22(3), 165–184. https://doi.org/10.2165/00019053-200422030-00003

Lohaus, D., & Habermann, W. (2019). Presenteeism: A review and research directions. Human Resource Management Review, 29(1), 43–58. https://doi.org/10.1016/j.hrmr.2018.02.010

Louviere, J. J., Flynn, T. N., & Marley, A. A. J. (2015). Best-Worst Scaling: Theory, Methods and Applications. Cambridge University Press. https://doi.org/10.1017/CBO9781107337855

MacLeod, C. M. (1991). Half a century of reseach on the stroop effect: An integrative review. Psychological Bulletin, 109(2), 163–203. https://doi.org/10.1037/0033-2909.109.2.163

MacNaughton, P., Satish, U., Laurent, J. G. C., Flanigan, S., Vallarino, J., Coull, B., Spengler, J. D., & Allen, J. G. (2017). The impact of working in a green certified building on cognitive function and health. Building and Environment, 114, 178–186. https://doi.org/10.1016/j.buildenv.2016.11.041

Mækelæ, M. J., Reggev, N., Dutra, N., Tamayo, R. M., Silva-Sobrinho, R. A., Klevjer, K., & Pfuhl, G. (2020). Perceived efficacy of COVID-19 restrictions, reactions and their impact on mental health during the early phase of the outbreak in six countries. Royal Society Open Science, 7(8), 200644. https://doi.org/10.1098/rsos.200644

Mamiseishvili, K., & Rosser, V. J. (2011). Examining the relationship between faculty productivity and job satisfaction. Journal of the Professoriate, 5(2).

Manzo, L. K. C., & Minello, A. (2020). Mothers, childcare duties, and remote working under COVID-19 lockdown in Italy: Cultivating communities of care. Dialogues in Human Geography, 10(2), 120–123. https://doi.org/10.1177/2043820620934268

Marley, A. A. J., & Louviere, J. J. (2005). Some probabilistic models of best, worst, and best–worst choices. Journal of Mathematical Psychology, 49(6), 464–480. https://doi.org/10.1016/j.jmp.2005.05.003

Martínez-Marquina, A., Niederle, M., & Vespa, E. (2019). Failures in contingent reasoning: The role of uncertainty. American Economic Review, 109(10), 3437–3474. https://doi.org/10.1257/aer.20171764

Martín-Fernández, J., Medina-Palomino, H. J., Ariza-Cardiel, G., Polentinos-Castro, E., & Rutkowski, A. (2018). Health condition and risk attitude in the Dutch population: An exploratory approach. Health, Risk & Society, 20(3–4), 126–146. https://doi.org/10.1080/13698575.2018.1458976

Maslach, C., & Jackson, S. E. (1981). The measurement of experienced burnout. Journal of Organizational Behavior, 2(2), 99–113. https://doi.org/10.1002/job.4030020205

Maslowski, R. (2020). Onderwijs (De Sociale Staat van Nederland). https://digitaal.scp.nl/ssn2020/onderwijs.

Mattke, S., Balakrishnan, A., Bergamo, G., & Newberry, S. J. (2007). A Review of Methods to Measure Health-related Productivity Loss. American Journal of Managed Care, 13(4), 7.

McDonald, J. H. (2014). Handbook of Biological Statistics. Sparky House Publishing.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. Frontiers in Econometrics.

McNeese-Smith, D. (1996). Increasing employee productivity, job satisfaction, and organizational commitment. Journal of Healthcare Management, 41(2), 160.

Merenda, P. F. (1997). A guide to the proper use of factor analysis in the conduct and reporting of research: Pitfalls to avoid. Measurement and Evaluation in Counseling and Development, 30(3), 156–164.

Mertens, G., Gerritsen, L., Duijndam, S., Salemink, E., & Engelhard, I. M. (2020). Fear of the coronavirus (COVID-19): Predictors in an online study conducted in March 2020. Journal of Anxiety Disorders, 74, 102258. https://doi.org/10.1016/j.janxdis.2020.102258

Meyer, A., Zhou, E., & Frederick, S. (2018). The non-effects of repeated exposure to the cognitive reflection test. Judgment and Decision Making, 13(3).

Miller, K. I., & Monge, P. R. (1986). Participation, satisfaction, and productivity: A meta-analytic review. Academy of Management Journal, 29(4), 727–753.

Moore, A. L. (2021, November 16). Keep it chill this holiday with new tools from Google Maps [Blog]. Google. https://blog.google/products/maps/keep-it-chill-holiday-new-tools-google-maps/

Morikawa, M. (2022). Work-from-home productivity during the COVID-19 pandemic: Evidence from Japan. Economic Inquiry, 60(2), 508–527. https://doi.org/10.1111/ecin.13056

Morse, S. S., Garwin, R. L., & Olsiewski, P. J. (2006). Next flu pandemic: What to do until the vaccine arrives? Science, 314(5801), 929–929.

Mosier, C. I. (1947). A critical examination of the concepts of face validity. Educational and Psychological Measurement, 7(2), 191–205.

Muchnik, L., Aral, S., & Taylor, S. J. (2013). Social influence bias: A randomized experiment. Science, 341(6146), 647–651. https://doi.org/10.1126/science.1240466

Mühlbacher, A. C., Kaczynski, A., Zweifel, P., & Johnson, F. R. (2016). Experimental measurement of preferences in health and healthcare using best-worst scaling: An overview. Health Economics Review, 6(1), 2. https://doi.org/10.1186/s13561-015-0079-x

Muraven, M., & Baumeister, R. F. (2000). Self-Regulation and Depletion of Limited Resources: Does Self-Control Resemble a Muscle? Psychological Bulletin, 126(2), 247–259. https://doi.org/10.1037/0033-2909.126.2.247

Nagel, R. (1995). Unraveling in guessing games: An experimental study. The American Economic Review, 85(5), 1313–1326.

Nakamura, M., Yoda, T., Crawshaw, L. I., Yasuhara, S., Saito, Y., Kasuga, M., Nagashima, K., & Kanosue, K. (2008). Regional differences in temperature sensation and thermal comfort in humans. Journal of Applied Physiology, 105(6), 1897–1906.

National Center for Immunization and Respiratory Diseases (NCIRD), D. of V. D. (2020). Travel during the COVID-19 Pandemic. https://www.cdc.gov/coronavirus/2019-ncov/travelers/travel-during-covid19.html

Nelson, B. W., Pettitt, A., Flannery, J. E., & Allen, N. B. (2020). Rapid assessment of psychological and epidemiological correlates of COVID-19 concern, financial strain, and health-related behavior change in a large online sample. PloS One, 15(11), e0241990. https://doi.org/10.1371/journal.pone.0241990

Newman, M. E. J. (2002). Spread of epidemic disease on networks. Physical Review E, 66(1), 016128. https://doi.org/10.1103/PhysRevE.66.016128

News, B. (2020, May 21). Coronavirus: Alarm as crowds flock to European beaches.

O'Donnell, A., Wilson, L., Bosch, J. A., & Borrows, R. (2020). Life satisfaction and happiness in patients shielding from the COVID-19 global pandemic: A randomised controlled study of the 'mood as information' theory. PLOS ONE, 15(12), e0243278. https://doi.org/10.1371/journal.pone.0243278

Ogden, R. (2021). Distortions to the passage of time during England's second national lockdown: A role for depression. Plos One, 16(4), e0250412.

Ospina, M. B., Dennett, L., Waye, A., Jacobs, P., & Thompson, A. H. (2015). A systematic review of measurement properties of instruments assessing presenteeism. The American Journal of Managed Care, 21(2), e171-85.

Palacios, J., Eichholtz, P., & Kok, N. (2020). Moving to productivity: The benefits of healthy buildings. PLOS ONE, 15(8), e0236029. https://doi.org/10.1371/journal.pone.0236029

Palia, D., & Lichtenberg, F. (1999). Managerial ownership and firm performance: A re-examination using productivity measurement. Journal of Corporate Finance, 5(4), 323–339.

Park, Y., Fritz, C., & Jex, S. M. (2011). Relationships between work-home segmentation and psychological detachment from work: The role of communication technology use at home. Journal of Occupational Health Psychology, 16(4), 457.

Parsons, K. (2014). Human thermal environments: The effects of hot, moderate, andcold environments on human health,comfort, and performance. CRC press. https://doi.org/10.1201/b16750

Pastor-Satorras, R., Castellano, C., Van Mieghem, P., & Vespignani, A. (2015). Epidemic processes in complex networks. Reviews of Modern Physics, 87(3), 925.

Peijen, R., Harakeh, Z., Niks, I., Brouwer, C., Hooftman, W., & Oude Hengel, K. (2022). De effecten van langdurig thuiswerken gedurende de coronapandemie op de gezondheid (TNO 2022 R10335). Ministerie van Sociale Zaken en Werkgelegenheid (SZW).

Peperkoorn, L. S., Becker, D. V., Balliet, D., Columbus, S., Molho, C., & Van Lange, P. A. M. (2020). The prevalence of dyads in social life. PLOS ONE, 15(12), e0244188. https://doi.org/10.1371/journal.pone.0244188

Pérez-Lombard, L., Ortiz, J., & Pout, C. (2008). A review on buildings energy consumption information. Energy and Buildings, 40(3), 394–398. https://doi.org/10.1016/j.enbuild.2007.03.007

Perry, J. L., & Porter, L. W. (1982). Factors Affecting the Context for Motivation in Public Organizations. Academy of Management Review, 7(1), 89–98. https://doi.org/10.5465/amr.1982.4285475

Pinheiro, C. A. R. (2011). Social Network Analysis in Telecommunications. Wiley. https://books.google.nl/books?id=jP8zfL6yNGkC

Plohl, N., & Musil, B. (2021). Modeling compliance with COVID-19 prevention guidelines: The critical role of trust in science. Psychology, Health & Medicine, 26(1), 1–12. https://doi.org/10.1080/13548506.2020.1772988

Poelman, M. P., Gillebaart, M., Schlinkert, C., Dijkstra, S. C., Derksen, E., Mensink, F., Hermans, R. C. J., Aardening, P., de Ridder, D., & de Vet, E. (2021). Eating behavior and food purchases during the COVID-19 lockdown: A cross-sectional study among adults in the Netherlands. Appetite, 157, 105002. https://doi.org/10.1016/j.appet.2020.105002

Prasad, M., Wahlqvist, P., Shikiar, R., & Tina Shih, Y.-C. (2004). A Review of Self-Report Instruments Measuring Health-Related Work Productivity. PharmacoEconomics, 22(4), 225–244. https://doi.org/10.2165/00019053-200422040-00002

Przybylski, A. K., Murayama, K., Dehaan, C. R., & Gladwell, V. (2013). Motivational, emotional, and behavioral correlates of fear of missing out. Computers in Human Behavior, 29(4), 1841–1848. https://doi.org/10.1016/j.chb.2013.02.014

Puente-Díaz, R., & Cavazos-Arroyo, J. (2022). Feeling grateful to be optimistic: The influence of recalling special moments on feelings of gratitude and optimism during the COVID-19 pandemic. International Journal of Psychology, 57(3), 336–340.

Qian, D., & Li, O. (2020). The Relationship between Risk Event Involvement and Risk Perception during the COVID-19 Outbreak in China. Applied Psychology: Health and Well-Being, 12(4), 983–999. https://doi.org/10.1111/aphw.12219

Raphael, K. (1987). Recall bias: A proposal for assessment and control. International Journal of Epidemiology, 16(2), 167–170.

Rattray, J., Johnston, M., & Wildsmith, J. (2004). The intensive care experience: Development of the ICE questionnaire. Journal of Advanced Nursing, 47(1), 64–73.

Rattray, J., & Jones, M. C. (2007). Essential elements of questionnaire design and development. Journal of Clinical Nursing, 16(2), 234–243. https://doi.org/10.1111/j.1365-2702.2006.01573.x

RIVM. (2020, March 15). Nieuwe maatregelen tegen verspreiding coronavirus in Nederland | Nieuwsbericht | Rijksoverheid.nl. https://www.rijksoverheid.nl/actueel/nieuws/2020/03/12/nieuwe-maatregelen-tegen-verspreiding-coronavirus-in-nederland

Roskam, E. E., Van den Wollenberg, A. L., & Jansen, P. G. (1986). The Mokken scale: A critical discussion. Applied Psychological Measurement, 10(3), 265–277.

Ross, M. (1989). Relation of implicit theories to the construction of personal histories. Psychological Review, 96(2), 341.

Salehi, B., Cordero, M. I., & Sandi, C. (2010). Learning under stress: The inverted-U-shape function revisited. Learning & Memory, 17(10), 522–530.

Sardeshmukh, S. R., Sharma, D., & Golden, T. D. (2012). Impact of telework on exhaustion and job engagement: A job demands and job resources model. New Technology, Work and Employment, 27(3), 193–207. https://doi.org/10.1111/j.1468-005X.2012.00284.x

Satish, U., Mendell, M. J., Shekhar, K., Hotchi, T., Sullivan, D., Streufert, S., & Fisk, W. J. (2012). Is CO2 an indoor pollutant? Direct effects of low-to-moderate CO2 concentrations on human decision-making performance. Environmental Health Perspectives, 120(12), 1671–1677. https://doi.org/10.1289/ehp.1104789

Sauermann, J. (2016). Performance measures and worker productivity. IZA World of Labor. https://doi.org/10.15185/izawol.260

Savadori, L., & Lauriola, M. (2021). Risk Perception and Protective Behaviors During the Rise of the COVID-19 Outbreak in Italy. Frontiers in Psychology, 11. https://www.frontiersin.org/articles/10.3389/fpsyg.2020.577331

Schacter, D. L. (1999). The seven sins of memory: Insights from psychology and cognitive neuroscience. American Psychologist, 54(3), 182.

Schacter, D. L. (2022). Memory Sins in Applied Settings: What Kind of Progress? [Working Paper].

Schäfer, A., Wand, B. M., Lüdtke, K., Ehrenbrusthoff, K., & Schöttker-Königer, T. (2021). Validation and investigation of cross cultural equivalence of the Fremantle back awareness questionnaire-German version (FreBAQ-G). BMC Musculoskeletal Disorders, 22(1), 1–14.

Schneider, C. R., Dryhurst, S., Kerr, J., Freeman, A. L. J., Recchia, G., Spiegelhalter, D., & van der Linden, S. (2021). COVID-19 risk perception: A longitudinal analysis of its predictors and associations with health protective behaviours in the United Kingdom. Journal of Risk Research, 24(3–4), 294–313. https://doi.org/10.1080/13669877.2021.1890637

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation Coefficients: Appropriate Use and Interpretation. Anesthesia & Analgesia, 126(5), 1763–1768. https://doi.org/10.1213/ANE.0000000000002864

Schwartz, B. (2000). Self-determination: The tyranny of freedom. American Psychologist, 55(1), 79–88. https://doi.org/10.1037/0003-066X.55.1.79

Sent, E.-M. (2018). Rationality and bounded rationality: You can't have one without the other. The European Journal of the History of Economic Thought, 25(6), 1370–1386. https://doi.org/10.1080/09672567.2018.1523206

Serafini, G., Parmigiani, B., Amerio, A., Aguglia, A., Sher, L., & Amore, M. (2020). The psychological impact of COVID-19 on the mental health in the general population. QJM: An International Journal of Medicine, 113(8), 531–537. https://doi.org/10.1093/qjmed/hcaa201

Shao, W., & Hao, F. (2020). Confidence in political leaders can slant risk perceptions of COVID–19 in a highly polarized environment. Social Science & Medicine (1982), 261, 113235. https://doi.org/10.1016/j.socscimed.2020.113235

Shepperd, J. A., Carroll, P., Grace, J., & Terry, M. (2002). Exploring the causes of comparative optimism. Psychologica Belgica, 42(1–2), 65–98.

Shepperd, J. A., Waters, E., Weinstein, N. D., & Klein, W. M. P. (2015). A Primer on Unrealistic Optimism. Current Directions in Psychological Science, 24(3), 232–237. https://doi.org/10.1177/0963721414568341

Shibasaki, M., Namba, M., Oshiro, M., Kakigi, R., & Nakata, H. (2017). Suppression of cognitive function in hyperthermia; From the viewpoint of executive and inhibitive cognitive processing. Scientific Reports, 7. https://doi.org/10.1038/srep43528

Shikiar, R., Halpern, M. T., Rentz, A. M., & Khan, Z. M. (2004). Development of the Health and Work Questionnaire (HWQ): An instrument for assessing workplace productivity in relation to worker health. Work, 22(3), 219–229.

Shikiar, R., Rentz, A., Halpern, M., & Khan, Z. (2001). The health and work questionnaire (HWQ): An instrument for assessing workplace productivity in relation to worker health. Value in Health, 4(2), 181–181.

Siegrist, M., & Árvai, J. (2020). Risk Perception: Reflections on 40 Years of Research. Risk Analysis, 40(S1), 2191–2206. https://doi.org/10.1111/risa.13599

Simon, H. A. (1956). Rational choice and the structure of the environment. Psychological Review, 63(2), 129.

Simon, H. A. (1990). Bounded Rationality. In J. Eatwell, M. Milgate, & P. Newman (Eds.), Utility and Probability (pp. 15–18). Springer. https://doi.org/10.1007/978-1-349-20568-4_5

Singh, H., Motwani, J., & Kumar, A. (2000). A review and analysis of the state-of-the-art research on productivity measurement. Industrial Management & Data Systems.

Skirbekk, V. (2004). Age and individual productivity: A literature survey. Vienna Yearbook of Population Research, 133–153.

Smeets, P., Bauer, R., & Gneezy, U. (2015). Giving behavior of millionaires. Proceedings of the National Academy of Sciences, 112(34), 10641–10644. https://doi.org/10.1073/pnas.1507949112

Sobkow, A., Zaleskiewicz, T., Petrova, D., Garcia-Retamero, R., & Traczyk, J. (2020). Worry, Risk Perception, and Controllability Predict Intentions Toward COVID-19 Preventive Behaviors. Frontiers in Psychology, 11. https://www.frontiersin.org/articles/10.3389/fpsyg.2020.582720

Somsen, G. A., Rijn, C. van, Kooij, S., Bem, R. A., & Bonn, D. (2020). Small droplet aerosols in poorly ventilated spaces and SARS-CoV-2 transmission. The Lancet Respiratory Medicine, 8(7), 658–659. https://doi.org/10.1016/S2213-2600(20)30245-9

Stagnaro, M. N., Pennycook, G., & Rand, D. G. (2018). Performance on the cognitive reflection test is stable across time. Judgment and Decision Making, 13(3). https://doi.org/10.2139/ssrn.3115809

Stahl, D. O. (1993). Evolution of Smartn Players. Games and Economic Behavior, 5(4), 604–617. https://doi.org/10.1006/game.1993.1033

Standen, P., Daniels, K., & Lamond, D. (1999). The home as a workplace: Work–family interaction and psychological well-being in telework. Journal of Occupational Health Psychology, 4(4), 368.

StataCorp. (2017). Stata Statistical Software: Release 15.

Statistieken over het Coronavirus en COVID-19. (2021, May 4). AlleCijfers.nl. https://allecijfers.nl/nieuws/statistieken-over-het-corona-virus-en-covid19/

Stevens, J. P. (2012). Applied multivariate statistics for the social sciences. Routledge.

Stochl, J., Jones, P. B., & Croudace, T. J. (2012). Mokken scale analysis of mental health and well-being questionnaire item responses: A non-parametric IRT method in empirical research for applied health researchers. BMC Medical Research Methodology, 12(1), 1–16.

Stroom, M., Eichholtz, P., & Kok, N. (2021). Avoiding Crowded Places During COVID-19: Common Sense or a Complex Strategic Decision? Frontiers in Psychology, 12. https://www.frontiersin.org/articles/10.3389/fpsyg.2021.700640

Stroom, M., Kok, N., Strobel, M., & Eichholtz, P. M. A. (2021). Turning up the heat: The impact of indoor temperature on selected cognitive processes and the validity of self-report. Judgment and Decision Making, 16(3), 766–795.

Sun, K., Chen, J., & Viboud, C. (2020). Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: A population-level observational study. The Lancet Digital Health, 2(4), e201–e208.

Swisher, K. (2013, February 22). "Physically Together": Here's the Internal Yahoo No-Work-From-Home Memo for Remote Workers and Maybe More. AllThingsD. https://allthingsd.com/20130222/physically-together-heres-the-internal-yahoo-no-work-from-home-memo-which-extends-beyond-remote-workers/

Syverson, C. (2011). What determines productivity? Journal of Economic Literature, 49(2), 326–365.

Tahmasebi, F., Wang, Y., Cooper, E., Godoy Shimizu, D., Stamp, S., & Mumovic, D. (2021). Window operation behaviour and indoor air quality during lockdown: A monitoring-based simulation-assisted study in London. Building Services Engineering Research and Technology, 01436244211017786. https://doi.org/10.1177/01436244211017786

Thissen, D., Steinberg, L., & Kuang, D. (2002). Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. Journal of Educational and Behavioral Statistics, 27(1), 77–83. https://doi.org/10.3102/10769986027001077

Thorp, A. A., Kingwell, B. A., Owen, N., & Dunstan, D. W. (2014). Breaking up workplace sitting time with intermittent standing bouts improves fatigue and musculoskeletal discomfort in overweight/obese office workers. Occupational and Environmental Medicine, 71(11), 765–771.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. Memory and Cognition, 39(7), 1275–1289. https://doi.org/10.3758/s13421-011-0104-1

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. Thinking and Reasoning, 20(2), 147–168. https://doi.org/10.1080/13546783.2013.844729

Train, K. E. (2009). Discrete Choice Methods with Simulation. Cambridge University Press.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Tyler, J. M., & Rosier, J. G. (2009). Examining self-presentation as a motivational explanation for comparative optimism. Journal of Personality and Social Psychology, 97(4), 716.

van der Lippe, T., & Lippényi, Z. (2020). Beyond Formal Access: Organizational Context, Working From Home, and Work–Family Conflict of Men and Women in European Workplaces. Social Indicators Research, 151(2), 383–402. https://doi.org/10.1007/s11205-018-1993-1 van der Torre, A., & Steenbekkers, A. (2020). Bevolking, economie en leefomgeving (De sociale staat van Nederland: 2020). Sociaal en Cultureel Planbureau. https://digitaal.scp.nl/ssn2020/bevolking-economie-en-leefomgeving

Van Lange, P. A. M., Rinderu, M. I., & Bushman, B. J. (2017). Aggression and violence around the world: A model of CLimate, Aggression, and Self-control in Humans (CLASH). Behavioral and Brain Sciences, 40. https://doi.org/10.1017/S0140525X16000406

Van Ooijen, A. M. J., Van Marken Lichtenbelt, W. D., Van Steenhoven, A. A., & Westerterp, K. R. (2004). Seasonal changes in metabolic and temperature responses to cold air in humans. Physiology and Behavior, 82(2–3), 545–553. https://doi.org/10.1016/j.physbeh.2004.05.001

Ventresca, M., & Aleman, D. (2013). Evaluation of strategies to mitigate contagion spread using social network characteristics. Social Networks, 35(1), 75–88.

Vermeulen, W. J. V., & Hovens, J. (2006). Competing explanations for adopting energy innovations for new office buildings. Energy Policy, 34(17), 2719–2735. https://doi.org/10.1016/j.enpol.2005.04.009

Virick, M., DaSilva, N., & Arrington, K. (2010). Moderators of the curvilinear relation between extent of telecommuting and job and life satisfaction: The role of performance outcome orientation and worker type. Human Relations, 63(1), 137–154. https://doi.org/10.1177/0018726709349198

Vittersø, J., Akselsen, S., Evjemo, B., Julsrud, T. E., Yttri, B., & Bergvik, S. (2003). Impacts of Home-Based Telework on Quality of Life for Employees and Their Partners. Quantitative and Qualitative Results From a European Survey. Journal of Happiness Studies, 4(2), 201–233. https://doi.org/10.1023/A:1024490621548

Wageman, R., & Baker, G. (1997). Incentives and cooperation: The joint effects of task and reward interdependence on group performance. Journal of Organizational Behavior, 18(2), 139–158. https://doi.org/10.1002/(SICI)1099-1379(199703)18:2<139::AID-JOB791>3.0.CO;2-R

Walker, G., Kogut, B., & Shan, W. (1997). Social capital, structural holes and the formation of an industry network. Organization Science, 8(2), 109–125.

Wang, B., Liu, Y., Qian, J., & Parker, S. K. (2021). Achieving Effective Remote Working During the COVID-19 Pandemic: A Work Design Perspective. Applied Psychology, 70(1), 16–59. https://doi.org/10.1111/apps.12290

Wang, X. (2017). An Empirical Study of the Impacts of Ambient Temperature on Risk Taking. Psychology, 08(07), 1053–1062. https://doi.org/10.4236/psych.2017.87069

Weinstock, M. A., Colditz, G. A., Willet, W. C., Stampfer, M. J., Rosner, B., & Speizer, F. E. (1991). Recall (report) bias and reliability in the retrospective assessment of melanoma risk. American Journal of Epidemiology, 133(3), 240–245.

Wells, M., & Perrine, R. (2001). Critters in the cube farm: Perceived psychological and organizational effects of pets in the workplace. Journal of Occupational Health Psychology, 6(1), 81–87. https://doi.org/10.1037/1076-8998.6.1.81

Welsh, M. B., Burns, N. R., & Delfabbro, P. H. (2013). The Cognitive Reflection Test: How much more than Numerical Ability? 35th Annual Conference of the Cognitive Science Society, 35(35), 1587–1592.

Wilson, A. E., & Ross, M. (2000). The frequency of temporal-self and social comparisons in people's personal appraisals. Journal of Personality and Social Psychology, 78(5), 928.

Wilson, A. E., & Ross, M. (2003). The identity function of autobiographical memory: Time is on our side. Memory (Hove, England), 11(2), 137–149. https://doi.org/10.1080/741938210

Wong, D. P. (2004). The effects of indoor air quality on performance and productivity. Indoor Air, 14(s7), 92–101. https://doi.org/10.1111/j.1600-0668.2004.00278.x

World Health Organization. (2020). Coronavirus disease (COVID-19) advice for the public. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public

World Health Organization Writing Group. (2006). Nonpharmaceutical interventions for pandemic influenza, national and community measures. Emerging Infectious Diseases, 12(1), 88.

Wright, K. P., Hull, J. T., & Czeisler, C. A. (2002). Relationship between alertness, performance, and body temperature in humans. American Journal of Physiology-Regulatory, Integrative and Comparative Physiology, 283(6), R1370–R1377. https://doi.org/10.1152/ajpregu.00205.2002.-Body

Wyon, D. P. (1974). The effects of moderate heat stress on typewriting performance. Ergonomics, 17(3), 309–317. https://doi.org/10.1080/00140137408931356

Yang, L., Holtz, D., Jaffe, S., Suri, S., Sinha, S., Weston, J., Joyce, C., Shah, N., Sherman, K., Hecht, B., & Teevan, J. (2022). The effects of remote work on collaboration among information workers. Nature Human Behaviour, 6(1), Article 1. https://doi.org/10.1038/s41562-021-01196-4

Yeh, C.-H., Hartmann, M., & Langen, N. (2020). The Role of Trust in Explaining Food Choice: Combining Choice Experiment and Attribute Best-Worst Scaling. Foods, 9(1), E45. https://doi.org/10.3390/foods9010045

Yeomans, K. A., & Golder, P. A. (1982). The Guttman-Kaiser Criterion as a Predictor of the Number of Common Factors. Journal of the Royal Statistical Society. Series D (The Statistician), 31(3), 221–229. JSTOR. https://doi.org/10.2307/2987988

Zagreus, L., Huizenga, C., Arens, E., & Lehrer, D. (2004). Listening to the occupants: A Web-based indoor environmental quality survey. Indoor Air, 14 Suppl 8, 65–74. https://doi.org/10.1111/j.1600-0668.2004.00301.x

Zhang, D. C., Highhouse, S., & Rada, T. B. (2016). Explaining sex differences on the Cognitive Reflection Test. Personality and Individual Differences, 101, 425–427. https://doi.org/10.1016/j.paid.2016.06.034

Zhang, F., & De Dear, R. (2017). University students' cognitive performance under temperature cycles induced by direct load control events. Indoor Air, 27(1), 78–93. https://doi.org/10.1111/ina.12296

Zhang, F., de Dear, R., & Hancock, P. (2019). Effects of moderate thermal environments on cognitive performance: A multidisciplinary review. Applied Energy, 236, 760–777. https://doi.org/10.1016/j.apenergy.2018.12.005

Zhang, X., Wargocki, P., Lian, Z., & Thyregod, C. (2017). Effects of exposure to carbon dioxide and bioeffluents on perceived air quality, self-assessed acute health symptoms, and cognitive performance. Indoor Air, 27(1), 47–64. https://doi.org/10.1111/ina.12284

Zhong, Y., Liu, W., Lee, T.-Y., Zhao, H., & Ji, J. (2021). Risk perception, knowledge, information sources and emotional states among COVID-19 patients in Wuhan, China. Nursing Outlook, 69(1), 13–21. https://doi.org/10.1016/j.outlook.2020.08.005

Zito, M., Ingusci, E., Cortese, C. G., Giancaspro, M. L., Manuti, A., Molino, M., Signore, F., & Russo, V. (2021). Does the end justify the means? The role of organizational communication among work-from-home employees during the COVID-19 pandemic. International Journal of Environmental Research and Public Health, 18(8), 3933.

Zivin, J. G., Hsiang, S. M., & Neidell, M. (2018). Temperature and human capital in the short and long run. Journal of the Association of Environmental and Resource Economists, 5(1), 77–105. https://doi.org/10.1086/694177

Zivin, J. G., & Neidell, M. (2012). The impact of pollution on worker productivity. American Economic Review, 102(7), 3652–3673. https://doi.org/10.1257/aer.102.7.3652

About the author

Martijn Stroom was born on 15th of March, 1989, in Maastricht, the Netherlands. He attained a Bachelor of Education (2010), a Bachelor of Science in Psychology and Neuroscience (2013), and a Master of Science in Human Decision Science (2015).

In 2016, he initially started a Ph.D. in the Faculty of Health, Medicine, and Life Sciences which aimed to develop a decision aid to support people who intended to stop smoking. In 2017, his interests motivated him to join the Real Estate group within the Finance department under the supervision of Piet Eichholtz and Nils Kok. In the first year of his Ph.D., Martijn completed a variety of courses on sustainable finance and econometrics, equivalent to a Master's Degree.

Martijn's field of research is behavioral science. His research varies across multiple topics, with the same recurring theme: investigating the inaccuracy of human judgment, factors that unconsciously influence resulting decisions, and the unawareness of that influence. During his Ph.D. he has focused on healthy buildings and productivity, the behavioral impact of crisis on policy, as well as human's perception of social networks. He uses lab and online experiments to explore the behavioral side of many real-life, relevant, economic and societal problems.

Working around the pandemic restrictions, Martijn made (brief: <6 weeks) visits to NOVA School of Business and Economics (Lisbon, Portugal), Rotman School of Management (Toronto, Canada), and Harvard University (Cambridge, USA). He presented his work at international conferences such as Judgement and Decision-making (SJDM), AREAUE, and Indoor Air (ISIAQ). He taught varies courses at undergraduate and graduate level, most prominently Behavioral Finance.

During his Ph.D., Martijn has been the organizer of the HealthBuild conference and SBE Science Slams, both aimed at bridging research and society/ industry. Furthermore, he entered the public and policy climate label debate for the Dutch residential housing market. Arguing from a behavioral perspective, he repeatedly featured in national news outlets.

In 2022, Martijn continued his work within the Real Estate Department as post-doctoral researcher.