IN PRAISE *of* AN INQUISITIVE MIND

A Festschrift in Honor of Jüri Allik
on the Occasion of his 70th Birthday

Edited by

Anu Realo

UNIVERSITY OF TARTU Press

IN PRAISE OF AN INQUISITIVE MIND

A Festschrift in Honor of Jüri Allik
on the Occasion of his 70th Birthday

IN PRAISE *of* AN INQUISITIVE MIND

A Festschrift in Honor of Jüri Allik
on the Occasion of his 70th Birthday

Edited by

Anu Realo

UNIVERSITY OF TARTU
Press

UNIVERSITY OF TARTU
Press

Published in Estonia by the University of Tartu Press.
www.tyk.ee

Cover design by Kalle Paalits
Page layout and design by Aive Maasalu

Presented to Jüri Allik on March 2, 2019 in Tartu, Estonia.

ISBN 978-9949-77-997-0

CONTENTS

The List of Contributors 7
Foreword 9

PART I: REFLECTIONS AND RECOLLECTIONS **13**

1 *Young Man Allik* 15
Jaan Valsiner

2 *Between Moscow and Tartu: Reminiscences on How and Ruminations on Why* 19
Ehtibar N. Dzhafarov

3 *The First* 29
Talis Bachmann

PART II: THEORETICAL ESSAYS **39**

4 *Two Realities* 41
Endel Tulving

5 *Model for Controlling Functions* 45
Aavo Luuk

6 *Food for Thought* 57
Robert R. McCrae

7 *Social Desirability from a Pragmatic Point of View* 67
Kenn Konstabel

PART III: EMPIRICAL CONTRIBUTIONS **79**

8 *Outliers in Perceptual Averaging: Over- or Underweight?* 81
Aire Raidvee

9 *When Subjective is the Most Objective: A Comparison of Different Fatigue Measures in the Visual Mismatch Negativity (vMMN) Task* 91
Kairi Kreegipuu and Nele Põldver

10 *Can't Wait or Want it Now? Impulsivity Relates to the Immediacy rather than the Delay Sensitivity Aspect of Temporal Discounting* 105
Andero Uusberg, Uku Vainik, and Kairi Kreegipuu

11 *Towards Meaningful Comparisons of Personality in Large-Scale Cross-Cultural Studies* 123
Gwen Gardiner, Kyle Sauerberger, and David Funder

12 *Why Sometimes a Man is more like a Woman: Insights into the "Gender Paradox" of Psychological Sex Differences around the World* 141
David P. Schmitt

13 *Personality Loosely Mirrors the Geography of the World* 151
Uku Vainik, Liisi Kööts-Ausmees, and René Mõttus

14 *Fatalism and External Locus of Control Are Different Constructs, or Allik on Allikas (Allik is the Source) of Many Good Ideas* 163
Lawrence T. White and Danielle R. Blazek

15 *Allik—A Common Surname in Southern Estonia?* 175
Markku Verkasalo

THE LIST OF CONTRIBUTORS

The Editor

ANU REALO, Department of Psychology, University of Warwick, United Kingdom; Institute of Psychology, University of Tartu, Estonia

The Contributors

TALIS BACHMANN, School of Law, University of Tartu, Estonia

DANIELLE R. BLAZEK, Department of Psychology, Claremont Graduate University, USA

EHTIBAR N. DZHAFAROV, Department of Psychological Sciences, Purdue University, USA

DAVID FUNDER, Department of Psychology, University of California, Riverside, USA

GWEN GARDINER, Department of Psychology, University of California, Riverside, USA

KENN KONSTABEL, National Institute for Health Development, Estonia; Institute of Psychology, University of Tartu, Estonia; School of Natural Sciences and Health, University of Tallinn, Estonia

LIISI KÖÖTS-AUSMEES, Institute of Psychology, University of Tartu, Estonia

KAIRI KREEGIPUU, Institute of Psychology, University of Tartu, Estonia

AAVO LUUK, Institute of Psychology, University of Tartu, Estonia

ROBERT R. MCCRAE, Gloucester, Massachusetts, USA

RENÉ MÕTTUS, Department of Psychology, University of Edinburgh, United Kingdom; Institute of Psychology, University of Tartu, Estonia

NELE PÕLDVER, Institute of Psychology, University of Tartu, Estonia

AIRE RAIDVEE, Institute of Psychology, University of Tartu, Estonia; New York University Abu Dhabi, United Arab Emirates

KYLE SAUERBERGER, Department of Psychology, University of California, Riverside, USA

DAVID P. SCHMITT, Centre for Culture and Evolution, Brunel University London, United Kingdom

ENDEL TULVING, University of Toronto, Canada

ANDERO UUSBERG, Institute of Psychology, University of Tartu, Estonia

UKU VAINIK, Montreal Neurological Institute, McGill University, Canada; Institute of Psychology, University of Tartu, Estonia

JAAN VALSINER, Department of Communication and Psychology, Aalborg University, Denmark

MARKKU VERKASALO, Department of Psychology and Phonetics, University of Helsinki, Finland

LAWRENCE T. WHITE, Department of Psychology, Beloit College, USA

MEMBERS OF THE INTERNATIONAL SITUATIONS PROJECT

FOREWORD

In Praise of an Inquisitive Mind: A Festschrift in Honor of Jüri Allik on the Occasion of his 70th Birthday

The noblest pleasure is the joy of understanding
Leonardo Da Vinci

This Festschrift is a celebration of Professor Jüri Allik's successful and productive career in various fields of psychology, on the occasion of his 70th birthday. The chapters published in this Festschrift, written by some of Jüri's many friends, colleagues, and former PhD students, reflect Jüri's diverse avenues of research and the various ways in which his insights and contributions have influenced the work of psychologists the world over. The contributions vary from more personal recollections and reminiscences (Part I) to theoretical reflections (Part II) and empirical studies (Part III). I hope that Jüri, as well as all other readers, will find the contributions to this volume both informative and stimulating.

Professor Jüri Allik is a rare species among scientists, a quintessential Renaissance man, whose talents and interests range across different fields of science—from psychophysics and visual perception to language processing, personality and cross-cultural psychology, history of psychology, and bibliometrics. The driving force behind this has been the ever-inquisitive mind of Jüri, which restlessly tackles new challenges and strives to answer the big questions. Sometimes, the answers occur to him in his sleep, which perhaps, at least partly, explains his success and effectiveness—his mind never seems to stop working and, as a result, by the time most people are slowly waking up and getting ready for the day, Jüri has already solved another fundamental psychological or statistical problem! So, it comes as no surprise that during his long and fruitful career, spanning more than four decades, Jüri has published more than 300 high-quality scientific papers, mostly in peer-reviewed international journals. Therefore, the sheer number of publications is already ample testimony to his remarkable productivity as a scientist. Yet, what is even more important is the spread and impact of his work. As of January 2019, these works have been cited over 18,000 times in Google Scholar, resulting in an *h* index of 58. These impressive figures speak for themselves and show no sign of abating. Quite on the contrary, the number of publications as well as citations has steadily increased over the past 20 years and is still on the growth curve.

In any field of science, as in life, it is far too easy to get lost in the details and to lose focus of what we are really trying to do or achieve. What makes Jüri and his research so special is his ability to see the bigger picture, his courage to ask the big questions and to

search for universal principles to explain human behavior and mental processes. Why are sex differences in personality larger in more advanced and developed societies? How can we explain the counterintuitive finding that there is more (and not less) social capital in more individualistic societies? Are personality traits geographically clustered? Is the way people see others different from how people see themselves? How does people's ability to recognize facial expressions of emotions change with age? These are just a few examples of the type of problems Jüri and his colleagues' research has tried to solve over the past two decades.

There are many discoveries in different fields of psychology to which Jüri has substantially contributed. In 1991, for instance, he and Tiia Tuulmets proposed the Occupancy Theory, which explained how the perceived numerosity of randomly distributed visual objects is formed in human perception. In 2005, Jüri and Kenn Konstabel published a paper in which they proved that Georges Frédéric Parrot, not Herman von Helmholtz, was first to propose the theory of unconscious inferences, which has served as a foundation for understanding human perception until today. One of the most surprising discoveries by Jüri, David Schmitt, and colleagues is perhaps their study published in 2008, in which they showed that sex differences in personality do not disappear with societal human development, but rather increase as societies become more affluent and egalitarian. Jüri's world-leading research on cross-cultural personality variation, together with Robert R. McCrae, René Mõttus, and many other colleagues, has revealed that cross-cultural differences are surprisingly small in magnitude and often do not concur with national stereotypes. Regardless of the topic or research field, the defining features of Jüri's work are scholarly rigor, methodological sophistication, and broad intellectual appeal.

During his career, Jüri has been a part of numerous international collaborations, often initiating them. Importantly, Jüri was involved in international collaborative research projects even when this was neither the norm nor easy in the then-Soviet Estonia. Indeed, Jüri's formative years were spent under the Soviet regime and he received his first PhD degree at the Moscow State University in 1976. He is living proof that the oppressive totalitarian regime—despite its best efforts—could not break the freedom of the mind or the autonomous thinking of its people. It is exactly Jüri's quest for knowledge and progress that allowed him to succeed under vast amounts of pressure both during the years of the Communist regime and in the early turbulent years of the Estonian re-independence in 1990s. Jüri was the first president and a founding member of the Union of the Estonian Psychologists (*Eesti Psühholoogide Liit*), which was established in 1988, and it is greatly thanks to Jüri's vision, commitment, and openness that psychological science in Estonia is probably more successful than in many other ex-communist countries. Most importantly, it is Jüri's uncompromising emphasis on research excellence and international collaboration that laid the foundation for the breakthrough of Estonian psychology into the "big league," as we know it today.

Jüri's association with the University of Tartu started in 1968 when he arrived in Tartu as a first-year psychology student. As an innovator and a generator of ideas, Jüri has made a significant contribution, and has already left an enduring legacy, to his Alma Mater, in his role as a professor, a Dean, a member of the Senate, and, particularly, as probably the longest-serving Head in the history of the Department of Psychology (later Institute). Over 20 PhD students and countless MSc and BSc students have successfully graduated from the University of Tartu under Jüri's supervision. Many of his former PhD students have become productive and internationally well-known researchers of their own standing, including eight of the authors in this Festschrift, and I believe that many of them will agree with me when I say that a significant part of their career success is likely due to Jüri's inspiring creativity, his insightful advice and strive for excellence.

In sum, honoring Jüri with a Festschrift does not really do justice to Jüri's contribution to the advancement of psychological science or to his extraordinary pivotal role in the development of the discipline in Estonia. Nevertheless, with this humble attempt, friends and colleagues in Estonia and all around the world wish to acknowledge the exceptional leadership and scholarship he has offered to the field of psychology, and their contributions are testimony to the extraordinary impact he has had on his many students, colleagues and co-workers, his department, the University of Tartu, and Estonian society in general. All of us who have been lucky enough to work with him have benefited from his infectious enthusiasm for new ideas, his wisdom and generosity, and from his ability to see and understand the "big picture," be it of a research field, a research funding policy, or of life in general.

In preparing this Festschrift, I have benefited from the kind assistance of several colleagues, friends, and family members. In particular, I would like to thank Liisi Kööts-Ausmees and Delaney Michael Skerrett for their most helpful editorial suggestions and assistance, and of course all the authors who kindly contributed to this book. I have greatly enjoyed editing this Festschrift and I was deeply touched by the effort and thoughtfulness put into each of the chapters. I am also grateful to the University of Tartu Press for agreeing to publish this work and to the Faculty of Social Sciences of the University of Tartu and its Dean, Professor Raul Eamets, for financial support. Last but not least, I would like to thank my daughter Annamari for her continuous support and encouragement throughout the editing process, but also for keeping the idea of the Festschrift a secret from her dear father. I can only hope that Jüri is pleased with the outcome and that all contributors and readers join me in wishing him a very happy 70th birthday and many exciting intellectual puzzles yet to be solved!

With love and appreciation,
Anu Realo

PART I:
REFLECTIONS AND RECOLLECTIONS

1 YOUNG MAN ALLIK

JAAN VALSINER

In this short essay, I analyze the history of a psychologist who has—in the Estonian context—become emblematic, a "hyper-generalized sign," of the advancement of science throughout the entire country. Here, I recall my joint efforts with Jüri Allik in the 1970s and point to the continuity of his youthful sentiments in the decades after that.

Birthdays are always moments for breaking out of year-long lying about our age and facing the truth for that one day. There is, in fact, an hour, even a minute, on that day, that is the precise time when real age and stated age match, but we easily forget such desire for precision. Thinking of birthdays is a good demonstration of the "hare and tortoise" classic paradox—the athlete who can run fast cannot overtake the tortoise—who is always infinitesimally ahead of the hare at any given moment. When Jüri Allik can honestly say "I am 70," it will be the last moment of truth—at least for a while. He will be lying for the next 364 days (until he can honestly say "I am 71"). In reality, he is every day becoming a day older than his stated age. It is probably not coincidental that one English way of wishing someone a happy birthday is to say "many happy returns" of the day. The truth will thereby prevail. By writing this short essay, I join all the other authors of this book in our best wishes for many happy returns of this Truth Day!

But truth can also seem untrue. When I met Jüri again—after some years—in April 2018 in Tartu, I realized that years have not really changed his ever ironically positive and inquisitive mind. In particular, I recollected an occasion during my first visit to Tartu in the Spring of 1970 in which I was aiming to find out if a timid post-adolescent like me (back then we did not think of such youngsters in terms of emerging adulthood) could somehow find a way to psychology after having been incredibly bored during my studies to become an English teacher. After attending a conference of the student scientific society and listening to many clever—and for me, intellectually overwhelming— presentations I found Jüri—then a 2nd year psychology student—lying down on somebody else's bed in a student dormitory ("Pälsoni inter") giving his appraisal of the event ("they are all stupid … and so are we!").

This was an active search for ways to go beyond the information given—penetrating through the presentational prose of ostensibly clever people to find the intellectual core of their thinking. Or perhaps to find that there is not one there. In psychology, the latter has continued to be the case—and I remain a humble disciple of Jüri Allik in my efforts to bypass the shining heights of currently popular fashions in

E-mail: jvalsiner@hum.aau.dk

the field—be it cognitivism or neuroscience. All these mainstream fashions continue to be superficial—and if we follow them—we may suffer a similar fate.

The young student Allik was the informal leader in the formation of psychology in Estonia in the 1970s. The Psychology Department at Tartu University was fertile ground for new—for Estonia—perspectives, and Jüri was leading the activities of a group of young psychology students, who, in their youthful enthusiasm, knew they were indeed building something new. There was freedom of scientific thought—paired with an absence of resources—and a deep desire to advance the field of psychology further. All this led to innovative business deals to get access to new technologies—operating within the shadow economy of the then-USSR. We became well trained in the real capitalism that operated behind official economic planning. It was no surprise that, after 1991, Estonia was well-prepared for the disappearance of the socialist economy, and the experience of dealing with Moscow was soon substituted with successful negotiations with Brussels.

We were intellectually cocky—no topic seemed to resist our deeply penetrating analysis. Sitting together with Jüri in an Aeroflot airplane to Moscow, we began planning an article on the role of the unconscious in the work of R. D. Laing. Immediately after reaching the ground, we did indeed write the article (Allik & Valsiner, 1977–1979, still unpublished), in which we claim that Laing had no concept of the unconscious at all. The finding might not seem very profound, but it is a good example of the prevailing ethos of intellectual omnipotence in young men within their self-invented intellectual milieu. It is therefore not surprising that a real milieu of such kind was created after Estonia re-gained political independence in 1991.

There are many possible images of a psychologist. René Magritte, back in 1948, had a rather romantic image of the power of the discipline:

Figure 1.1 "Psychologist" by René Magritte (1948). The Musée Magritte Museum, Royal Museums of Fine Arts of Belgium.

Psychologists in Estonia might not all live up to this powerful image, which resulted from Magritte's only encounter with psychologists (see Figure 1.1), but they have developed power within their field that has certainly found an adequate place within society. Intellectually, the psychologists who emerged in Estonia in the period 1974 to 1991 were already free in their professional lives—as the Soviet system had utterly failed to set up any role for psychologists within society. Once these roles did get established in the 1990s, psychologists became not only specialists but—in a semiotic sense—signs (Kullasepp, 2019). Their presence within a society indicates a level of maturity. In a semiotically savvy society like Estonia, this kind of generalization is not unexpected.

Jüri Allik as a Sign

If psychologists are signs, and Jüri Allik is a psychologist, he must also be a sign. But he cannot be a simple kind of sign. In fact, in the last three decades, he has become a good example of a *hyper-generalized sign* (Cabell & Valsiner, 2011). His guidance of psychology through society holistically captures the general atmosphere of the entire field and directs its future development. Allik's guidance of the field in independent Estonia is well recognized by all who understand the ways in which a discipline develops in a small but ambitious country. There are probably more personality researchers per capita in Estonia than anywhere in the world. After Allik moved into personality psychology in the early 1990s, Estonia became the leading nation in Europe in that field of psychological science. The only other comparable areas where such success has taken place through the semiotic facilitation of a field is in classical music, where the legacies of Arvo Pärt and Neeme Järvi have set Estonia up as *the* place for innovations in classical music.

The Young Man at 70

Jüri Allik remains young in his productively sarcastic yet positive ways of being. He lives up to his credo—to always avoid boring people and situations (Allik, 2017). Of course, in the academic context, this is no easy task, as we all easily succumb to the dubious pleasures of debates in various committees and of public polemics about the socially visible landscapes of science-society relations. Jüri's youthful enthusiasm in bringing about change within the developing Estonian society can be seen if one looks at those public disputes he decides to participate in, and the many others that he lets blissfully pass by. The youthful ethos of the 1970s is still there and actively at work, decades later. So, it barely matters that one birthday passes, again—the mind keeps moving ahead.

References

Allik, J. (2017). *Väldi igavaid inimesi ja olukordi [Avoid boring people and situations]*. Tartu: University of Tartu Press. [In Estonian].

Allik, J., & Valsiner, J. (1977–1979). *The concept of unconscious in the work of R. D. Laing.* Unpublished manuscript.

Cabell, K. R., & Valsiner, J. (2011). Affective hypergeneralization: Learning from psychoanalysis. In S. Salvatore & T. Zittoun (Eds.), *Cultural psychology and psychoanalysis: Pathways to synthesis* (pp. 87–113). Charlotte, NC: Information Age Publishers.

Kullasepp, K. (2019). *Presenting oneself as a psychologist: How social representation processes guide professional development*. London: Routledge.

2 BETWEEN MOSCOW AND TARTU: REMINISCENCES ON HOW AND RUMINATIONS ON WHY

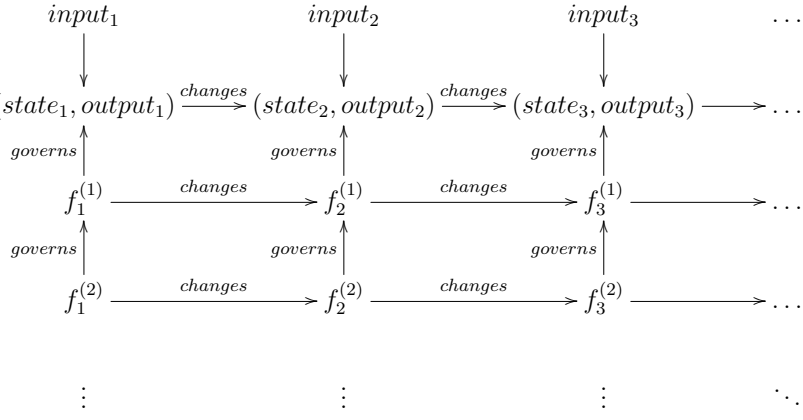
EHTIBAR N. DZHAFAROV

A glimpse into the history of two young men who were intent on doing "normal science" in the hostile environment of Soviet psychology.

When written by a long-time friend, it is not unusual for a Festschrift paper to be a hodgepodge of reminiscences: from scientific ideas related to the honoree's work to anecdotes related to the honoree's life. If my arithmetic is correct, I have known Jüri Allik for just a couple of years shy of half a century, so I do qualify as a long-time friend. In fact, Jüri is my oldest friend with whom I still communicate and share interests.

When we first met, in the early 1970s, Jüri was primarily interested in the semiotics of cultural phenomena, such as folklore. I was a sophomore at Moscow University, and my ambition was to create a general mathematical theory of sentient behavior. We did not make a great match, quite obviously. Our interest in each other, however, grew very fast, and I attribute this to three interrelated reasons: our shared hatred of the communist regime, our contempt for Soviet-style psychology, and our determination to do "normal," good-quality science. I will talk of the former two later—first, the science.

During my first year of knowing Jüri, I talked to him of nothing else but the "hierarchy of machines" as a universal language in which sentient behavior could be described. Jüri may find it amusing to recall this construction now.



E-mail: ehtibar@purdue.edu

A machine of the first order is a function $f_t^{(1)}$ that takes an input and a system's state at time t, and maps them into some observable output and a new state at the next moment:

$$f_t^{(1)} : (state_t, input_t) \rightarrow (state_{t+1}, output_{t+1}).$$

The idea is that this function itself can be viewed as a state of a second-order machine (without observable outputs):

$$f_t^{(2)} : (f_t^{(1)}, input_t) \rightarrow f_{t+1}^{(1)};$$

and that the second-order machines form the states of a third-order machine, and so on:

$$f_t^{(3)} : (f_t^{(2)}, input_t) \rightarrow f_{t+1}^{(2)},$$

$$f_t^{(4)} : (f_t^{(3)}, input_t) \rightarrow f_{t+1}^{(3)},$$

etc.

Sometimes I was saying "up to some unknown topmost level," but the possibility of an infinite hierarchy was especially fascinating, because it was combining complete determinism with the complete unpredictability of observable outputs based on inputs and current states. The last time I mentioned this construction to someone, having translated it into the hierarchy of differential equations, it was to Lee Rubel, a very good mathematician at the University of Illinois at Urbana-Champaign. Lee, regrettably, passed away shortly afterwards.

Jüri, as we know, successfully survived my expounding all this, perhaps because, at least in the beginning, he was not listening. He got interested, however, in my other crazy idea, and this is how our collaboration began. I was thinking of apparent motion, a very popular subject at the time, and it struck me that since the image of the motion can only begin after the second flash's onset, the physical duration of this image need not be the same as the physical interval between the onsets of the two flashes. If so, I reasoned, what if this image does not change in time at all? What if its content (the image of a light spot moving from one location to another) is, in its entirety, present at any moment of physical time for as long as the image lasts? I told this to Jüri, and we designed an experiment in which an observer had to press a key as soon as the image of the moving spot passed a visual marker (a small line) placed just under the motion's apparent trajectory. The position of the marker randomly varied between a place close to the first flash and a place close to the second flash, and the expectation was that the difference between the mean response times for these two positions, measured from the onset of the second flash, would be zero.

Jüri programmed the experiment on a peculiar paper-tape-controlled computer-like device controlling LEDs, and his students and colleagues served as participants. For several days I conducted this experiment in Jüri's lab on Tiigi Street, sometimes staying there late in the night. We computed the arithmetic means of response times, and the difference was indeed statistically zero. I rushed to inform of this groundbreaking outcome my advisor, Aleksei Leontiev, who at the time was the official head of Soviet psychology. Leontiev was excited, so much that he mentioned this result in his next lecture. It would have been the beginning of glorious days for us if Jüri did not call me (or come to Moscow to tell me?) that the experiment, either the data recording or the analysis, I don't recall which now, was irredeemably bungled and belonged in a trash can. I think this experiment was conducted by others later on, with the result that the difference in response times is the same as it would be for real motion. It is a testament to my radically changed interests that I continue to write this without being curious enough to do an internet search to find out.

After the flop with "unchanging images of change," Jüri and I did quite a bit of serious work. Thus, we studied the ability of observers to identify the location of an apparent motion in a two-alternative forced choice paradigm, and to conjointly identify the direction of the motion. The seemingly paradoxical result obtained by Jüri was that the probability of a correct direction identification was higher than the probability of a correct detection of motion's location. This could, however, be readily explained by a Thurstonian-type decision making model based on sensory effects X_1 and X_2 of the two observation intervals. Assume, e.g., that X_1 and X_2 are independent, normally distributed random variables with the same variance; the mean of X_i (i = 1, 2) is 0 if the stimulus in the ith observation interval is stationary, otherwise the sign of the mean indicates the direction of motion. The decision rule for the direction identification then can be

$$\textit{"choose rightward motion if and only if } X_1 + X_2 > 0\textit{,"}$$

whereas the decision rule for the detection can be

$$\textit{"choose interval 1 if and only if } (X_1)^2 - (X_2)^2 > 0\textit{."}$$

Denoting the probability of a correct direction identification by p, it is easy to show then that this value must exceed $\frac{1}{2}$, and the probability of a correct detection is

$$p^2 + (1 - p)^2 < p.$$

The p-vs-$[p^2 + (1 - p)^2]$ graph did, in fact, describe the data quite accurately (Dzhafarov & Allik, 1980). There is no paradox, as one can compute that the conditional probability of a correct direction identification given an incorrect detection judgment

is $\frac{1}{2}$. The conjoint detection-identification scheme was later expanded by Jüri to several other types of stimuli, such as dot locations and luminance increments and decrements (Allik, Dzhafarov, & Rauk, 1982). The situation with some of these was more complex, leading us to speculate about different kinds of stimulus continua, those with "left zero" and those with "middle zero," as well as try more sophisticated models of sensory effects.

Another, very long line of research dealt with motion direction identification in what we first called "cinematograms" and then "kinematograms," to make them less suggestive of going to movies (Allik & Dzhafarov, 1984a). This was a sequence of regular patterns of elements each of which could be in one of two states. Each subsequent frame was created by shifting or rotating the previous one by a fixed amount in one of two opposite directions, and then reversing the state of each element with some fixed probability. A natural model accounting for the psychometric functions here is one comparing imperfect counts of the dipoles (pairs of the same-state elements in different frames) pointing in opposite directions. We later conducted the same type of research with the same type of models on other types of stimuli: e.g., we studied identification of "melody direction" (ascending or descending) in a sequence of random musical cords (Allik, Dzhafarov, & Ross, 1984; Allik, Dzhafarov, Houtsma, Ross, & Versfeld, 1989). And in 1984, during my month-long visit at the Bulgarian Academy of Sciences, I used a mathematically similar design and similar modeling to study apparent tilt of random dot patterns within circular apertures.

In retrospect, all this work was good-quality but rather unimaginative: it involved no "big ideas." We came closer to the latter around 1980, in our studies of motion detection as a function of the time-position function of a single dot, line, or any other rigid pattern. At that time motion detection in visual psychophysics was primarily operationalized as one's ability to identify motion direction in a unidirectionally moving low-contrast grating. The high-contrast sharp-edged stimuli of older motion perception research were in disrepute. It was a bold move from us therefore to return to them. It was a lucky move too, because we had no technical means of measuring contrast thresholds. Most of my work at the time, some of it in collaboration with Jüri, consisted in taking a perceptual phenomenon obtained by measuring contrast thresholds, such as selective adaptation to grating frequency, and constructing its experimental analogue using geometric or kinematic thresholds of high-contrast stimuli (Dzhafarov & Allik, 1981a, 1981b).

There was a loosely guiding principle behind all this research, a variant of James Gibson's "psychophysical" approach to perception. I formulated it in my 1979 PhD thesis (Dzhafarov, 1979): to understand a phenomenon means to compute a physical property of stimulus that is responsible for this phenomenon (e.g., one that has to reach a particular magnitude for the phenomenon to occur). The only difference from Gibson's approach was that the phenomenon in question was understood as pertaining to a specific judgment required of an observer (e.g., the task of detecting whether a dot was moving) rather than to perceptual images *per se*. Different tasks could be associated with different physical properties of one and the same stimulus.

For the experiments on motion detection I was able to obtain a light-beam oscilloscope that allowed us to create various oscillatory motions without worrying about electrochemical processes at the stimulus source. The dependent variable was the minimal amplitude of motion at a given frequency, or the maximal frequency of motion at a given amplitude, at which the oscillatory motion appeared different from a stationary object. After we had published our first paper on the topic, faithfully documenting differences in the thresholds for sine-wave, triangle-wave, and square-wave horizontal oscillations (Dzhafarov, Allik, Linde, & Pyastolov, 1981), we developed a model in which the relevant property of the time-position function $x\,(t)$ was its variance within a sliding temporal window $[t-\tau, t]$,

$$V\,(t) = \frac{1}{2\tau} \int\int_{u,v\in[t-\tau,t]} (x\,(u) - x\,(v))^2\,\mathrm{d}u\mathrm{d}v.$$

It was this variance $V\,(t)$ (or its mean over some interval of time) that had to reach a critical value for the motion $x\,(t)$ to become discernible from a stationary stimulus (Dzhafarov & Allik, 1984). Many years later, I discussed this model with George Sperling, and pointed out to him that his motion detection model with Charlie Chubb (at the time considered state of the art) would not be able to predict anything in such experiments, and in fact could not handle oscillatory motions at all. So he knows.

We expanded our model later, using some published data found by Jüri, to account for reaction times to onset of unidirectional motions (Allik & Dzhafarov, 1984b). Later, in the United States, I worked for a couple of months in Robert Sekuler's lab at Northwestern, and we collected data on reaction time to motion velocity changes, also well accounted for by our sliding variance model (Dzhafarov, Sekuler, & Allik, 1993).

The model itself is by no means a "big idea." There was, however, a genuine "big idea" suggested by this model. The computation of $V\,(t)$ was over a finite interval $[t-\tau, t]$, but its result was assigned to a point in time, t. That is, $V\,(t)$ was a local, instantaneous property of motion, the same as its position or velocity, but it was computed from an extended chunk of motion. The generalization of this way of looking at our model was dubbed by us the "global-in-local" principle: in perceptual judgments, all stimulus features are computed over extended spatiotemporal areas of stimulus, but the results of these computations (mathematically, functionals) are assigned to points in space and time. For instance, given a stimulus $L\,(x, y)$, a receptive field $W\,(x, y)$ computes a number

$$\mathrm{M} = \int\int_{(x,\,y)\,\in\,\mathrm{R}} L\,(x, y)\; W\,(x, y)\;\mathrm{d}x\mathrm{d}y.$$

over the receptive field's area R. Mathematically, by appropriately defining $W\,(x, y)$, one could even consider this R to be the entire visual field. However, the number M, in accordance with the "global-in-local" principle, should be assigned to a particular point, e.g., the center $(\bar{x}_R, \bar{y}_R)$ of the receptive field:

$$M = M(\bar{x}_R, \bar{y}_R).$$

Given a set of receptive fields whose centers densely cover the visual field, this will translate the "raw" stimulus $L(x, y)$ into a processed (or pre-processed, in view of subsequent decision making) stimulus $M(x, y)$.

I continue to think to this day that the "global-in-local" principle is a correct way of analyzing perception: a percept is a superposition of all pre-processed stimuli like $M(x, y, t)$, each computed by a different set of functionals each of which gathers information from some spatiotemporal area and assigns it to a point in space and time. I am not sure if we ever published a formulation of this principle. I remember discussing it with Narendra Ahuja in the late 1980s, at the Beckman Institute in Champaign, Illinois. He was, as it seemed to me, genuinely impressed by the principle as a way of developing machine vision, but somehow, we never got to work on this in earnest. In the early 2000s, I ran into Narendra at an airport, both of us stranded overnight by bad weather, and the first thing he asked me was if I remembered this principle. I did.

The photograph in Figure 2.1 was probably taken shortly before my leaving Moscow for Prague, in 1982. I had stayed in that Tartu home many times before, sleeping on a sofa in Jüri's study, taking long walks with him during the day, and having long late-night talks in his kitchen with its big wood-burning stove. We had at the time several projects for future collaboration, and for the period of my life in Prague, we continued to collaborate at a distance. Many results of this collaboration were lost later, after I "defected to the West" in 1986. Thus, I lost the results of my 1984 experiments on tilt perception, mentioned earlier. I also lost the Master's thesis and forgot the name of a talented young student I advised in Prague: the thesis contained a very general Thurstonian-type model of the conjoint detection and recognition, with sensory effects multivariate-normally distributed.

Following my "defection," our contacts with Jüri had to be severed, for obvious reasons. Unfortunately, our collaboration was never restored, even after the monstrous entity called the Soviet Union had blissfully disappeared and Jüri found himself in a free Estonia. Our last joint publication was Dzhafarov, Sekuler, and Allik (1993), on the reaction time to motion velocity changes, but it was based on much older results. Jüri visited with me at both my Illinois and Indiana places of work, I visited him in Tallinn and Tartu, we have done a great deal of discussing, but new collaboration (at least as of today) has not materialized.

The main reason for this was that in the 1990s I had largely lost interest in the "normal" work of a perceptual psychologist: conducting experiments and fitting narrowly-aimed models to the results. In my recent and current research, I have been much closer to the freshman constructing hierarchies of machines for understanding sentient behavior than to the young researcher constructing and testing models for random kinematograms. In retrospect, the latter never was my cup of tea. It was highly instructive for me, no doubt, and the research we did with Jüri was good science, but

Figure 2.1 Jüri and his son Alo playing soccer, with me watching, in the backyard of his old home in Tartu. This was around 1982, possibly my last visit to Tartu before my emigration to then-Czechoslovakia.

I recall our chats during those long walks in snow-clad Tartu with greater fondness than most of the papers we published together. I am writing this with no deprecation in mind: my lack of enthusiasm for something does not belittle it. Anyone who knows my work in sensory discrimination functions (e.g., Dzhafarov, 2003a; Dzhafarov & Colonius, 2006; Dzhafarov & Dzhafarov, 2012), generalized Fechnerian Scaling (Dzhafarov & Colonius, 1999, 2007), selective influences (Dzhafarov, 2003b; Kujala & Dzhafarov, 2008), and quantum contextuality (Dzhafarov, Cervantes, & Kujala, 2017; Dzhafarov & Kujala, 2012, 2017; Kujala, Dzhafarov, & Larsson, 2015) will see that my enthusiasm at present belongs in a different kind of research. Jüri, too, has drifted into a completely new area, personality theory, although he retains to this day some of his interest in modeling perceptual judgments.

There are several explanations for why in those youthful years I chose to do and enjoyed doing the work I did in collaboration with Jüri. One was in fact mentioned earlier, as a reason for Jüri and me being drawn together despite our very different backgrounds and initial interests. This reason was our shared contempt for Soviet-style psychology, which included not only the official doctrines "in the light" of which all Soviet psychologists were doing their things, but also the various sects and semi-forbidden semi-Marxist thinkers who were attracting throngs of admiring intelligentsia. As a reaction to their "dialectical" blabbering and Byzantine disputes, we were naturally drawn to what we considered good, honest science, as it was known to

us from the hard-to-obtain international journals of experimental psychology: conduct experiments, fit models, compare to others' models. Even outside the Soviet Union this would have been a rational choice at earlier stages of one's career. Without the discipline brought in by this restricted view of science, even with good mastery of rigorous mathematics, one could easily confuse elucidating mathematical theorizing with all-obscuring mathematical metaphorizing. Johann Herbart was a good example of the latter (see Dzhafarov, 2006), as opposed to Gustav Fechner and, even clearer, Duncan Luce, who exemplified the former.

A closely related explanation is that experimental work with narrowly defined math-intensive modeling provided some degree of defense against ideological tests and prohibitions. Jüri and I could publish our papers with no references to Marx, Lenin, or Lev Vygotsky, and we could even write our papers in English and submit them to Western journals. By contrast, my first published paper was theoretical, on mathematical modeling in psychology (Leontiev & Dzhafarov, 1973), and my mentor Aleksei Leontiev found it necessary to replace in it a simple remark to the effect that a percept was not a model of the perceived with a reference to "the standard Marxist view," according to which a subjective image was "a transformed, existence-in-the-other form of an object." Freedom from ideology, unfortunately, did not imply freedom from patriotism, even in very technical publications. Thus, in my paper on the relationship between detectability and reaction time (Dzhafarov, 1977), Leontiev insisted that I begin it with a reference to "domestic" (literally, "fatherland") psychophysics.

(Interestingly enough, Soviet-style psychology seems to have been fully restored in the academia of modern Russia. The post-Soviet psychology in that country has mirrored in its development the path undergone by the Russian state as a whole, from the initial realization of inferiority to the eventual plunging with abandonment back to the cozy parochialism of "our special path." Russia, currently a major source of chaos in the world, has branded both its people and its science with hot iron.)

This brings me to the last of the three reasons for why Jüri and I were drawn together in those youthful years: the abhorrence of Soviet communism that we immediately recognized in each other. My negative feelings were so intense that it became progressively harder for me to camouflage them. Even the mastered-by-all playacting in order to pass the university exams in the ideological disciplines became an almost unsurmountable obstacle for me by the time I was about to graduate. I was lucky to be under the protection of my mentor Aleksei Leontiev: I may have graduated without it, but not without much more grief. Jüri did not have a powerful protector, but he had the advantage of living in Estonia, where strong nationalist sentiments made it easier for a person to be a hater of the regime. To be able to discuss, condemn, and ridicule the regime in our conversations was as important to us as our science, and to me at least, a way of preserving sanity.

It is worth noting that nationalism (or patriotism) is one aspect of worldview in which we may have had differences with Jüri. To me, patriotism is as evil as communism or fascism, and I have felt so since my adolescence. I have systematically taught myself to never have feelings of pride or shame about things related to where I was born or live. Fittingly, I have kept emigrating (and with the present-day monstrosity in the White House, I would emigrate again if I knew where to). Jüri remained in his country and saw it recover from the communist stranglehold and flourish. I think he has genuine loyalty to Estonian culture and pride in its successes, to which he has himself contributed by his work and his prominence. And this is one case where I am willing to make an exception and suppress my disapproval. After all, it is not wrong to like a country one rationally considers good, even if it happens to be the country of one's birth.

References

Allik, J., & Dzhafarov, E. N. (1984a). Motion direction identification in random cinematograms: A general model. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 378–393.

Allik, J., & Dzhafarov, E. N. (1984b). Reaction time to motion onset: Local dispersion model analysis. *Vision Research, 24,* 99–101.

Allik, J., Dzhafarov, E. N., Houtsma, A. J. M., Ross, J., & Versfeld, N. J. (1989). Pitch motion with random chord sequences. *Perception & Psychophysics, 46,* 513–527.

Allik, J., Dzhafarov, E. N., & Rauk, M. (1982). Position discrimination may be better than detection. *Vision Research, 22,* 1079–1081.

Allik, J., Dzhafarov, E. N., & Ross, J. (1984). The identification of melody direction in random chord sequences. In M. Rauk (Ed.), *Computational models in hearing and vision* (pp. 5–9). Tallinn: Estonian Academy of Sciences.

Dzhafarov, E. N. (1977). The structure of the process of detection. *Vestnik MGU: Seriya Psikhologia [MGU Journal: Series of Psychology], 4,* 52–65. [In Russian].

Dzhafarov, E. N. (1979). *Elementary constituents of visual perception.* Moscow: Moscow State University Press. (Summary of the unpublished PhD thesis). [In Russian].

Dzhafarov, E. N. (2003a). Thurstonian-type representations for "same-different" discriminations: Deterministic decisions and independent images. *Journal of Mathematical Psychology, 47,* 208–228.

Dzhafarov, E. N. (2003b). Selective influence through conditional independence. *Psychometrika, 68,* 7–26.

Dzhafarov, E. N. (2006). Mathematical theorizing versus mathematical metaphorizing: A commentary on Rudolph. *Culture and Psychology, 12,* 205–214.

Dzhafarov, E. N., & Allik, J. (1980). "Sign-side paradox" for apparent motion detection. *Proceedings of the XXII International Congress of Psychology, 2,* 615. Leipzig.

Dzhafarov, E. N., & Allik, J. (1981a). Multiple-feature structures in psychophysical tasks. In A. Kashin (Ed.), *Problems of psychophysics and differential psychophysiology* (pp. 33–55). Kazan: Kazan University Press. [In Russian].

Dzhafarov, E. N., & Allik, J. (1981b). Selective visual adaptation and the visual feature model. In J. M. Zabrodin (Ed.), *Psychological studies in perception and memory* (pp. 77–92). Moscow: Nauka. [In Russian].

Dzhafarov, E. N., & Allik, J. (1984). A general theory of motion detection. In M. Rauk (Ed.), *Computational models in hearing and vision* (pp. 77–84). Tallinn: Estonian Academy of Sciences.

Dzhafarov, E. N., Allik, J., Linde, N. D., & Pyastolov, V. K. (1981). Comparative analysis of the frequency-amplitude threshold functions for real and apparent motion. *Psikhologicheskii Zhurnal [Journal of Psychology], 2,* 73–78. [In Russian].

Dzhafarov, E. N., Cervantes, V. H., & Kujala, J. V. (2017). Contextuality in canonical systems of random variables. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 375*(2106), 20160389. doi:10.1098/rsta.2016.0389

Dzhafarov, E. N., & Colonius, H. (1999). Fechnerian metrics in unidimensional and multi-dimensional stimulus spaces. *Psychonomic Bulletin and Review, 6,* 239–268.

Dzhafarov, E. N., & Colonius, H. (2006). Regular minimality: A fundamental law of discrimination. In H. Colonius & E. N. Dzhafarov (Eds.), *Measurement and representation of sensations* (pp. 1–46). Mahwah, NJ: Erlbaum.

Dzhafarov, E. N., & Colonius, H. (2007). Dissimilarity cumulation theory and subjective metrics. *Journal of Mathematical Psychology, 51,* 290–304.

Dzhafarov, E. N., & Dzhafarov, D. D. (2012). The sorites paradox: A behavioral approach. In L. Rudolph (Ed.) *Qualitative mathematics for the social sciences: Mathematical models for research on cultural dynamics* (pp. 105–136). London: Routledge.

Dzhafarov, E. N. & Kujala, J. V. (2012). Selectivity in probabilistic causality: Where psychology runs into quantum physics. *Journal of Mathematical Psychology, 56,* 54–63.

Dzhafarov, E. N., & Kujala, J. V. (2017). Probabilistic foundations of contextuality. *Fortschritte der Physik [Progress in Physics], 65, 6–8, 1*6000040. doi:10.1002/prop.201600040

Dzhafarov, E. N., Sekuler, R., & Allik, J. (1993). Detection of changes in speed and direction of motion: Reaction time analysis. *Perception and Psychophysics, 54,* 733–750.

Kujala, J. V., & Dzhafarov, E. N. (2008). Testing for selectivity in the dependence of random variables on external factors. *Journal of Mathematical Psychology, 52,* 128–144.

Kujala, J. V., Dzhafarov, E. N., & Larsson, J.-Å. (2015). Necessary and sufficient conditions for extended noncontextuality in a broad class of quantum mechanical systems. *Physical Review Letters, 115,* 150401. doi:10.1103/PhysRevLett.115.150401

Leontiev, A. N., & Dzhafarov, E. N. (1973). Mathematical modeling in psychology. *Voprosy psikhologii [Psychological Problems], 3,* 3–14. [In Russian].

3 THE FIRST

TALIS BACHMANN

Internationally visible peer-reviewed articles are the main 'luggage' of scientists when travelling along their career paths. Among the Estonian psychologists residing in Estonia, Jüri Allik has the highest citation rate for the research papers he has (co-)authored. For the scientific psychology community, both young and established, it may be of some interest to learn about his first international peer-reviewed article, which was published in *Perception* in 1976. What follows is an account of that article and circumstances around it.

Introduction

Almost everything unfolding throughout time has its benchmarks, both objective and indisputable, as well as conventional. For example, the year 1879 is known as the year the science of psychology was born, thanks to Wilhelm Wundt's lab being established in Leipzig. Yet, this, by no means, means that experimental psychology actually started in that year. Wundt himself, his many students and a number of other scholars like Müller, Helmholtz, and Fechner had already prepared the stage in their earlier studies and research. But it is useful to have some specific date to mark the beginning of something, even though it is not the most precise of all possible truths. If not for anything else, then at least for celebrative or student textbook purposes, it is nice to have such a date. Here, I take the liberty to comment on the first international peer-reviewed experimental research paper published by Estonian psychologists after the Second World War. It is understandably related to the scholar who gave the reason for the present Festschrift. Jüri Allik is not only the most cited local colleague (18,772 citations according to Google Scholar and 8,911 according to Elsevier's Scopus, both as at December 6, 2018), but he has been also the main and most influential proponent of the central role of scientometric statistical data on publications and citations in scientific evaluation (e.g., Allik, 2003). Jüri's career as a scientist began, of course, before the year 1976 (e.g., Allik, 1974), but the article published in *Perception* in 1976, in collaboration with the author of this essay, might be considered as some sort of benchmark.

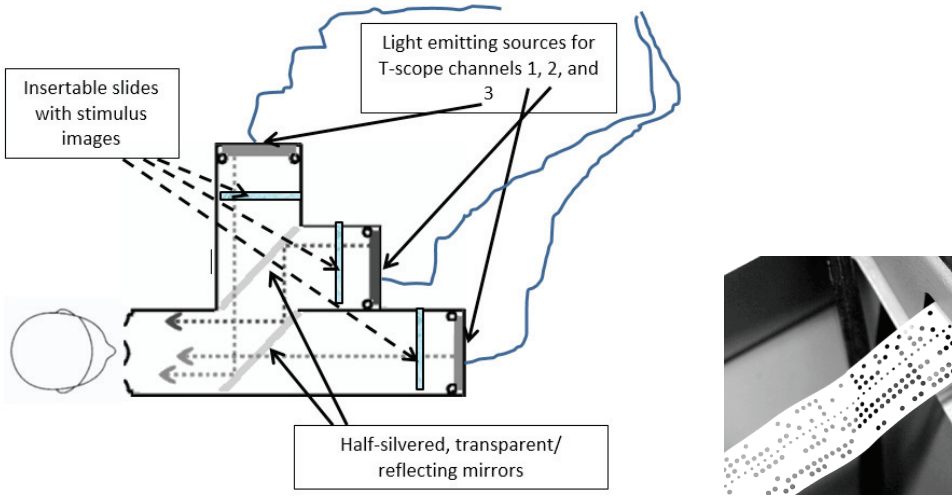
E-mail: talis.bachmann@ut.ee

The Premises and Origins of the Paper

In the mid-1970s a group of young Estonian psychologists, graduates of the University of Tartu (then Tartu State University), became fascinated by experimental studies in visual perception. The topics included eye movements, motion perception, and visual masking. The group included Jüri Allik, Aavo Luuk, and Talis Bachmann. One of the reasons we became 'visionaries' is easy to pinpoint: our Department of Psychology hosted professor Vladimir Petrovich Zinchenko (1931–2014) who, in 1973, gave a short series of lectures on perception that were eye-opening for some sitting in the audience. Although a professor from Moscow Lomonossov University, Zinchenko gave an account not only on his own creative research on eye movements and stabilized visual images, but also of contemporary perception research worldwide. This covered not only psychophysics but also an overview of the paradigm-changing information-processing approach (e.g., Haber, 1969; Neisser, 1967). Among the several positive results of this scientific contact was the willingness of Vladimir Petrovich to supervise the candidate (equivalent of PhD) thesis of Jüri Allik and MSc and candidate theses of Talis Bachmann (who became a graduate student of Zinchenko from 1974 to 1977). Despite the fact that our scientific supervisor was in Moscow, the experimental equipment necessary to 'revolutionize' Estonian experimental psychology was built in Tartu and assembled at 78 Tiigi Street, where the psychology department was located at the time (Figure 3.1).

Figure 3.1 Building of 78 Tiigi Street in Tartu, the home of the Department of Psychology at the University of Tartu from 1973 to 2013. The visual perception laboratory was based in the rooms hidden behind the façade in the wing of the building.

As every student of visual perception knows, in order to conduct reliable and valid visual perception experiments, presentation times and inter-stimulus intervals of visual stimuli must be controlled with temporal precision at the millisecond (ms) scale. The traditional means for this purpose consisted of tachistoscopic experimental equipment. Tachistoscopes (T-scopes) allow the presenting of visual flashes and images precisely, for durations of a few ms up to hundreds and thousands of ms. However, if experimental designs require that several stimuli have to be presented for the observer in rapid succession, and that time intervals between the stimuli must be precisely preset and measured, a multichannel T-scope is needed. (Historically, the best-known and most used prototypes were Gerbrands and Scientific Prototype apparatuses). As apparent motion, saccadic suppression, and visual masking research presuppose tachistoscopic methods, a T-scope was in urgent need in Tartu. The old mechanical T-scopes designed according to the Wundtian tradition and inherited from the Konstantin Ramul's (1879–1975) time in Tartu were most certainly outdated. Under the initiative of young psychologists Jüri Allik, Aavo Luuk, and Jaan Huik, and thanks to the skills of Psychology Department engineers Väino Vaske and Mihkel Miil, as well as Vahur Tuberg from the experimental mechanics workshop of the University of Tartu, the required apparatus was ready for use in 1974. A three-channel optical set was built by Vahur Tuberg (mimicking the typical design used worldwide; see Figure 3.2). This allows presenting up to three different spatially overlapping images for the observer. From the laboratory supervised by Arved-Aleksander Tammik, neighboring Vahur Tuberg's workshop, light emitting sources (luminophores) and half-silvered semi-transparent mirrors were obtained and inserted into the T-scope. The toughest problem—precise timing of the light pulses emitted from the luminophores, serving as the background light for depicting the dark images of the slides inserted into the T-scope—was solved elegantly by 'reinstating' the data feed module of the old Soviet computer, Ural 4. A punched tape was prepared so that the minimum distance between the perforations for each channel allowed a 5-ms temporal resolution between successive light flashes. When the tape was run in the Ural data feed module, precisely timed signals were sent to the purpose-built device that controlled the timing (switch-on and switch-off moments) of the stimuli in the T-scope optical module. Experimenters were able to prepare stimulation timing for their own experiments by setting the necessary distance of perforation between neighboring holes in the same perforation row of the tape. (A little metallic piercer used for this is vividly in the visual long-term memory of the present author).



A B

Figure 3.2 Panel A: Schematic depiction of the tachistoscopic setup for visual experiments; top view. The T-scope box is dark inside, and stimuli are visible only when light from the light source (luminophore) travels through the stimulus slide on which the darker than background image are drawn. Panel B: Example of a perforated tape similar to the one used for T-scope timing by experimental psychologists in the Department of Psychology at the University of Tartu in the early 1970s.

The study that developed into that first international peer-reviewed publication by Bachmann and Allik (1976) consisted of two experiments on visual masking. Two brief spatially overlapping and mutually different visual object forms (see panel A in Figure 3.3) were presented successively. In the first experiment, observers had to identify both objects. As a result, two functions of identification were obtained, showing the dependence of correct perception on the time interval between the stimulus objects. At the shortest interval, both objects, 1st and 2nd, were perceived at a comparable level of accuracy—above chance, but not perfectly. With intermediate delays between the objects, the second one prevailed in accurate perception, depriving the first one of clear perception, as if replacing the first with the second. Only with delays approaching 150 ms were both objects clearly perceived as two successive flashes of distinct objects. In the second experiment, the task was changed: observers had to detect the presence of the form, the identity of which was announced before each trial. In this experiment, the strong non-monotonic function of masking for the 1st object disappeared, and the target object was quite clearly perceived, regardless of its temporal position. Figure 3.3 (panel B) shows the results of this mutual masking study. (Mutual masking means that object 2 serves as the backward mask for object 1 and object 1 serves as the forward mask for object 2). By virtue of this design, several questions can be answered about the emergence of visual perception: how long does it take for a visual small object to reach conscious perception (answer—about 100–150 ms); whether a

succeeding briefly flashed object can replace the preceding briefly flashed object in consciousness (answer—yes, it can and does this best with delays between objects onsets equal to about 40–100 ms); whether selective attention tuned to expect a certain object can facilitate perception of that object, despite being covered in space by a different, masking object (answer—yes, it can). The results of this study anticipated and continually inform several current disputes in perception research, but let me comment on this a bit later.

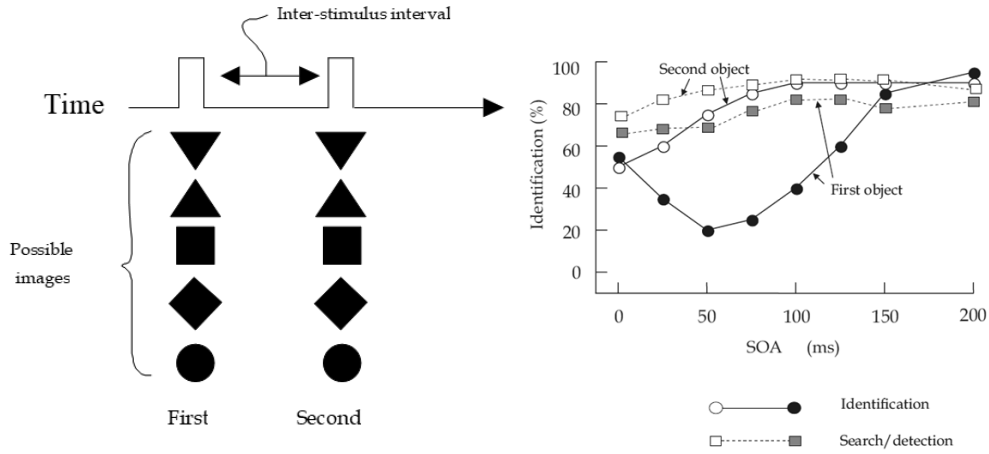


Figure 3.3 Panel A: Examples of visual forms presented, in the study by Bachmann and Allik (1976), as successive brief objects from the overlapping spatial location. The two forms presented in each trial were always mutually different. Panel B: Functions of masking obtained in Experiment 1 (identification of both stimuli) and Experiment 2 (detection of the pre-defined target in the search task).

A curious remark: one of the participants in these experiments observing the flashes in the T-scope was Endel Põder, the former student and now best-known Estonian researcher in visual search and visual crowding (e.g., Põder, 2017; Põder & Wagemans, 2007). Perhaps this experience from his student days was one of the reasons why he chose a scientific psychology career.

The Road to Publication

The results of this study were written up as a manuscript and suddenly the bold idea emerged to try to publish it in one of the main international journals specializing in perception. At that time, the main specialized publication outlets for visual perception research were *Perception and Psychophysics* (Psychonomic Society), *Vision Research* (Pergamon Press, now part of Elsevier) and *Perception* (Pion). In 1975, when the manuscript was ready, Talis Bachmann was already a graduate student of Zinchenko. A couple of years earlier, Richard Gregory (of cognitive theory of perception fame, see Gregory, 1966), who was the Editor-in-Chief for *Perception*, had invited Zinchenko to join the Editorial Board, an invitation that was accepted by Vladimir Petrovich. It seemed justified and logical to try that journal. This was not, it should be noted, an initiative of Zinchenko. At that time, very few and only well-established Soviet researchers had published in Western journals. So Zinchenko felt that he could not even consider doing this. When he ultimately came to know that the youngsters had submitted the paper to *Perception*, he seemed a bit surprised, but remained, of course, supportive.

The manuscript arrived at the editorial office on October 8, 1975. It was of course sent out for review by competent scholars. Curiously enough, when the message arrived at the turn of the year that it was not rejected, but a revision was requested, I was lucky enough to discover who the main reviewer was. As is typical, a blind review process was used, but as the reviewer kindly offered his help in improving the English of the paper, and had marked some questions and edits on the pages of the manuscript (sent back to the authors), I noticed that handwriting was somewhat familiar. (MS Word or rtf files for electronic editing were understandably unknown at that time). I then realized that the same hand had written some compliments on a reprint of a masking paper sent some time earlier upon my request. Yes, it was Max Coltheart, one of the leaders in cognitive psychology and also author of some seminal masking articles. Many years later, I had the opportunity to ask whether Max was indeed the reviewer and received an affirmative answer (with no demand to keep this fact confidential).

After we revised the text of the manuscript and extended the statistical analysis (ANOVA was supplemented by χ^2), it was sent back to the journal. I quite well remember one day in early 1976: when I arrived at 78 Tiigi Street and walked towards the lab along that long corridor, Jüri happened to open the door just at that very moment and, noticing me approaching, he could not contain his excitement so that he started jogging towards me with that naughty-boy smile on his face. The post had arrived earlier that morning with a positive response. One of the two co-authors was of course allowed to open it before the corresponding author arrived. The two-experiment study on mutual masking was now a reality (Bachmann & Allik, 1976). Jüri Allik himself has also commented on the birth of this article (Allik, 2017, p. 74 and p. 106), which in itself must constitute a relevant supplement to the present account.

Theoretical Context of the Paper

When I now read that paper, I again notice that it had several useful aspects and findings. First, it helped to show that effective masking can also be obtained between spatially overlapping successive, high-contrast visual objects, but in order to explain the masking effects involved in the inter-stimulus interaction, no single theory is sufficient. Both integration theory and interruption theory, complemented with the notion of attentional switch from object 1 to object 2, needed to be combined to understand the effects. Second, the paper showed that masking is much more closely related to attention than many had thought. Although object substitution masking theory (Di Lollo, Enns, & Rensink, 2000) is the best known and most influential visual masking theory conceptualized in terms of attention, in Bachmann and Allik (1976, p. 93) we may read the following:

> *When S1 is already represented at the iconic level, then this representation tends to trigger some other representation at the categorical level, i.e. it will be categorized or encoded. This search of category takes a certain amount of time… When the features of S2 are integrated before the encoding of S1 is completed, then the succeeding item replaces the 'old' icon with the representation of a new object. As the iconic level is already a conscious (or quasi-conscious) level, the subject must at the first opportunity, while continuing to encode S1 give, say, the name 'triangle' to the disc which he sees. But this inconsistent outcome is ruled out by the internal consistency of brain functioning, and the earlier conflicting response (or read-out process) is interrupted. The subject is unable to pay conscious attention to two objects at once, although they are represented at different levels. On the neurophysiological plane this is possibly done on the basis of inhibitory relationships between different more or less unitary representational systems, or by distortion of intercortical excitatory feedback loops.*

One can just notice and develop from this the backpropagation/reentrance idea! Even though reentrant processing theories (Di Lollo et al., 2000; Lamme & Roelfsema, 2000) are nowadays among the leading theories in trying to explain conscious perception and masking, some similar ideas in this context have also been expressed earlier. Thus, as the third notable aspect of that study, we can highlight the notion of reentrant processes. Furthermore, in the 60s and 70s of the 20th century, a debate was ongoing between the proponents of monotonic and those of non-monotonic functions of masking. From Bachmann and Allik (1976), it becomes clear that, in the same stimulation setup, both monotonic and nonmonotonic functions can be found, depending on the task and on the order of stimuli. This can be considered as the fourth important aspect of the article.

Aftermath and Coda

The paper on vision by Bachmann and Allik (1976) did not remain invisible. It has acquired about 80 references (Google Scholar) in specialist publications and has even been cited in various textbooks and handbooks on perception (e.g., Coren, Ward, & Enns, 1999; Uttal, 1981). Subsequent to the article, both Jüri and I continued experimental behavioral research on perceptual processes, but independently. He concentrated more on motion perception, eye movements, visual estimation of numerosity, and a couple of other topics. My research dealt with visual masking, selective attention, proactive visual facilitation effects, perception of pixelated images, and so on. While Allik remained more or less within behavioral psychophysics and abstract models (e.g., Allik, 1989; Allik & Dzhafarov, 1984; Allik & Tuulmets, 1991), I moved more and more towards psychophysiology of perceptual microgenesis and what is currently called cognitive neuroscience (e.g., Bachmann, 1984, 1991, 1994). I even remember that we once swapped books: I had Stevens' (1975) psychophysics and he had Haber's (1969) information-processing book, so we exchanged them for mutual benefit. (I still have the book on my bookshelf and perhaps Jüri has Stevens' on his). It may have been out of a wish to repay me for initiating our joint publication that Jüri later invited me to co-author a commentary article on iconic memory (Allik & Bachmann, 1983), which has remained to this day our last common international research publication. (There were also some co-authored forewords to local Estonian publications and a book review on Lloyd Kaufmann's *Sight and Mind* for a leading Russian psychology journal and of course the festive article on the occasion of the 80th birthday of our former mentor Vladimir Zinchenko; Allik, Bachmann, Luuk, & Tulviste, 2011). Then, about a quarter of century ago, Jüri's scientific interests turned heavily toward studying personality (e.g., Allik & McCrae, 2004; Rietveld et al., 2013; Schmitt, Allik, McCrae, & Benet-Martínez, 2007). When I asked him about this change, he mentioned that papers on personality tend to receive several times more citations than papers on perception. Even until now, I do not know whether it was said tongue-in-cheek. My own research at the same time became increasingly more focused on the problems of the neural correlates of consciousness (e.g., Aru, Bachmann, Singer, & Melloni, 2012; Bachmann, 2000; Phillips, Bachmann, & Storm, 2018). Nevertheless, I am sure Jüri will agree with me that our sweetest memories as up-and-coming scientists are of the time we spent in that good old 78 Tiigi Street building in the mid-seventies.

References

Allik, J. (1974). Существование и взаимодействие детекторов в зрите льном восприятия человека (по материалам зарубежных работ) I [On the existence of detectors and their interrelations in human visual perception I]. In M. Kotik (Ed.), *Труды по психологии III [Studies in Psychology III]*, *Acta et Commentationes Universitatis Tartuensis, 335*, 128–149. [In Russian, abstracts in Estonian and English].

Allik, J. (1989). Is a unified psychophysical law realistic? *Behavioral and Brain Sciences*, *12*, 267–268.

Allik, J. (2003). The quality of science in Estonia, Latvia, and Lithuania after the first decade of independence. *TRAMES: A Journal of the Humanities & Social Sciences*, *7*, 40–52.

Allik, J. (2017). *Väldi igavaid inimesi ja olukordi [Avoid boring people and situations]*. Tartu: University of Tartu Press. [In Estonian].

Allik, J., & Bachmann, T. (1983). How bad is the icon? *Behavioral and Brain Sciences, 35*, 12–13.

Allik, J., Bachmann, T., Luuk, A., & Tulviste, P. (2011). Vladimir Petrovich Zinchenko 80's birthday. *Kul'turno-istoricheskaya psikhologiya [Cultural-Historical Psychology], 3,* 3–4. [In Russian, abstracts in English].

Allik, J., & Dzhafarov, E. N. (1984). Motion direction identification in random cinematograms: A general model. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 378–393.

Allik, J., & McCrae, R. R. (2004). Toward a geography of personality traits: Patterns of profiles across 36 cultures. *Journal of Cross-Cultural Psychology*, *35*, 13–28.

Allik, J., & Tuulmets, T. (1991). Occupancy model of perceived numerosity. *Perception & Psychophysics*, *49*, 303–314.

Aru, J., Bachmann, T., Singer, W., & Melloni, L. (2012). Distilling the neural correlates of consciousness. *Neuroscience and Biobehavioral Reviews*, 36, 737–746.

Bachmann, T. (1984). The process of perceptual retouch: Nonspecific afferent activation dynamics in explaining visual masking. *Perception and Psychophysics, 35*, 69–84.

Bachmann, T. (1991). Identification of spatially quantised tachistoscopic images of faces: How many pixels does it take to carry identity? *European Journal of Cognitive Psychology, 3*, 87–103.

Bachmann, T. (1994). *Psychophysiology of visual masking: The fine structure of consciousness*. Commack, New York: Nova Science Publishers.

Bachmann, T. (2000) *Microgenetic approach to the conscious mind*. Amsterdam/ Philadelphia: John Benjamins.

Bachmann, T., & Allik, J. (1976). Integration and interruption in the masking of form by form. *Perception, 5*, 79–97.

Coren, S., Ward, L. M., & Enns, J. T. (1999). *Sensation and perception*. Fort Worth, TX: Harcourt Brace.

Di Lollo V., Enns J. T., & Rensink R. A. (2000). Competition for consciousness among visual events: The psychophysics of reentrant visual processes. *Journal of Experimental Psychology: General, 129,* 481–507.

Gregory, R. L. (1966). *Eye and brain: The psychology of seeing*. London: Weidenfeld & Nicolson.

Haber, R. N. (Ed.) (1969). *Information-processing approaches to visual perception*. New York: Holt, Rinehart & Winston.

Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences, 23,* 571–579.

Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.

Phillips, W. A., Bachmann, T., & Storm, J. F. (2018). Apical function in neocortical pyramidal cells: A common pathway by which general anesthetics can affect mental state. *Frontiers in Neural Circuits*, *12*, 1–15.

Põder, E. (2017). Combining local and global limitations of visual search. *Journal of Vision*, *17(4)*:10, 1–12.

Põder, E., & Wagemans, J. (2007). Crowding with conjunctions of simple features. *Journal of Vision*, *7(2)*:23, 1–12.

Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. W., ... & Albrecht, E. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science*, *340*, 1467–1471.

Schmitt, D. P., Allik, J., McCrae, R. R., & Benet-Martínez, V. (2007). The geographic distribution of Big Five personality traits: Patterns and profiles of human self-description across 56 nations. *Journal of Cross-Cultural Psychology*, *38*, 173–212.

Stevens, S. S. (1975). *Psychophysics*. New York: Wiley.

Uttal, W. R. (1981). *A taxonomy of visual processes*. Hillsdale, MI: Erlbaum.

PART II: THEORETICAL ESSAYS

4 TWO REALITIES

ENDEL TULVING

Consciousness is a universal brain/mind capacity that allows the organism to become aware of, and to interact with, its internal and external environment. It is the foundation of mental reality, a reality postulated to exist along with conventional physical reality. This paper posits that there is nothing in physical reality that is not also in our mental reality. Mental reality also "matters" to us humans in that it affects every aspect of the world we live in; physical reality is less relevant and matters only to the extent that its constituents correspond to the constituents of mental reality.

Jüri Allik, my friend and soul mate, the renaissance man, does not like boring people or boring ideas. Therefore, in this Festschrift in his honor, it would be appropriate to write something that is not boring. To make sure of this, it would be even better if it were out of this world, or at least something that many people would not approve of.

This is my attempt at this.

I tell a story about consciousness and two realities. In the long history of the learned study of consciousness, almost everything that has been written is wrong in the opinion of other experts in the field. What I say here fits right in, since everybody knows that there is no such thing as a mental reality, and for me to claim that there is, is preposterous, which is as good as out of this world.

***

Scientists hardly ever ask questions about the existence of things. Existence is taken for granted. Asking if something exists is not important in order to pursue our mission, which is to study and increase our knowledge of the world. Grass, trees, and apples exist; our bodies, our clothes, our computers exist. Thousands upon thousands of other things exist. They are part of reality, and reality is physical.

Philosophers are more interested in pondering the existence of things, be they physical or not. They have had much more experience with handling difficult, sometimes unanswerable questions. Indeed, a branch of metaphysics is ontology, which deals with questions such as "What is existence?", "What is a thing?", "What is a physical object?"

For many people, physical reality is the only reality. Things exist only in physical reality. Even things that happen in people's minds are explained as "products" of the brain.

E-mail: tulving@psych.utoronto.ca

Psychologists in general do not agree with this state of affairs. There is more in this world to study and ask questions about than physical objects. There exists a huge, wonderful world in the human brain. It consists of thoughts, ideas, feelings, visions, memories, imaginations, and more. These are all real.

The notion that reality is all physical is the dominant stance in science as well as many laypeople's view of the world. The problem is that this view restricts what we can study about mental phenomena, what kinds of questions we can or should ask.

There is a solution to this problem. It is very simple. Given that thoughts exist, as we know they do, we will create a place for them—a new reality, a "mental reality," where mental phenomena can exist legitimately.

In fact, it turns out that you do not need to create one—it already exists, and I am simply pointing it out.

The general idea of a non-physical reality has been proposed and discussed by philosophers and other thinkers from the earliest days, under a variety of appellations. Thus, there is nothing new or startling about the essence of my proposal that there is something more than just physical reality.

Let us define some concepts.

"Reality" is the sum total of everything that exists in a given domain. This definition, just a slightly different wording from what you would find in the dictionary, should sound reasonable to all reasonable people.

"Physical reality" is made up of physical objects and energy forms. These are things that do not depend on a conscious mind for their existence. Examples are volcanoes, lakes, snow, apples, neurons, a human body.

"Mental reality" is the sum total of all mental processes, including but not limited to thoughts, emotions, imaginations, problem-solving, perceptions, intentions, memories. Mental reality is enabled by consciousness. Every person has a mental reality unique to them, but it has some elements in common with other people. Table 4.1 is a list of selected examples of items in mental reality. The items in the list convey a picture of the richness of human experience and knowledge.

"Consciousness" is a general brain-mind capacity that enables an organism to become aware of its internal and external environment and interact with it. Consciousness is the foundation for mental reality, in the sense that nothing happens in mental reality in the absence of consciousness.

Let us think about why we consider something to be part of mental reality. Take the well-known conundrum: A tree falls in a forest; does it make a sound if there is no one around to hear it? The correct answer to the question, of course, is no. Sound is made (enabled) by two things, a source, like a falling tree, AND a biological auditory system, like the one in the functioning human brain, that can perceive the sound waves and create a mental experience, which is a part of that organism's mental reality.

Table 4.1 A list of selected examples of things that would not exist if mental reality did not exist.

Reality	Observation	Truth	Wealth
Property	Memory	Information	Envy
Relation	Remembering	Ignorance	Jealousy
Change	Nothingness	Mathematics	Gratitude
Similarity	Repetition	Measurement	Help
Difference	Completeness	Government	Regret
Language	Emptiness	Organization	Ecstasy
The past	Vacuum	Education	Joy
The future	Consciousness	Literature	Anger
Infinity	Awareness	Religion	Evil
Recency	Attention	Finance	Extraversion
Latency	Thinking	Sports	Neurosis
Novelty	Thought	Intelligence	Heroism
Permanence	Knowledge	Beauty	…
Science	Expertise	Angels	…

When you start thinking about it, you find that the distinction between the two kinds of realities is quite clear. The existence, and hence the reality, of any thing either does or does not depend on the mind. There are no things that are "a little bit dependent on the mind," as there are no such states as "a little bit pregnant," or a "little bit dead". Please note that this definition does not depend on any particular conceptualization of the mind. No reasonable definition of the mind could upset the logic of this argument.

An easy way to describe the relationships that I have talked about is to make a diagram of them, similar to a Venn diagram (Figure 4.1). On the left side of the figure are two overlapping sets of items, namely the two realities: A being the physical and B being the mental reality. They are independent of each other but partially overlap. Let us look at region B on the left side of Figure 4.1. This represents all items in mental reality that have no "correspondents" in physical reality. Again, see the examples listed in Table 4.1. We could call it "pure" mental reality.

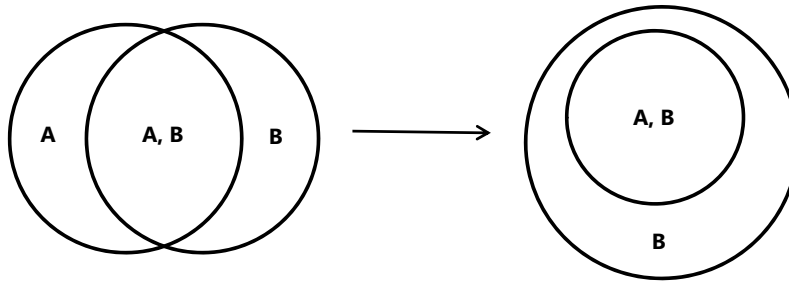


Figure 4.1 Venn diagram of the known universe. A indicates physical, and B mental, reality, respectively.

Next let us look at the intersection of A and B that contains both things from physical and mental reality, respectively. All items in this overlap are from physical reality and every one of them has a correspondence to at least one item in mental reality, and vice versa. Examples would be grasshopper, volcano, potato chip, golf ball.

Now consider region A on the left side of the figure. This represents all items in physical reality that do not have any correspondents in mental reality. Here comes a big surprise. Region A is empty. There is nothing in it. That means there are no things that exist in physical reality that humans are not aware of in some way or another. It is hard to imagine, but true. If you want to prove this assertion wrong, all you need to do it is to think of a single example that is a physical thing that humans are not aware of in some way. Good luck with this exercise!

It is surprising to realize that everything in physical reality is also part of your mental reality simply because you can perceive it or think about it. People may not like it, but it does not change the fact that there is no such thing as *pure* physical reality because all of our experience with physical reality is based on our mental reality.

If nothing exists in pure physical reality, then the diagram can be redrawn as on the right side of Figure 4.1. This shows that all of physical reality is encompassed by mental reality. The relationship that we observe in this second diagram could be interpreted almost as if mental reality is what allows for things to exist in physical reality. An item in physical reality has to have a match, or correspondent, in mental reality for it to exist. If it does not have a match, it does not exist.

The two realities have a tangled relationship. Mental reality is wholly dependent on physical in that, like all things alive, it is created by physical reality. Yet mental reality, as it has evolved over eons of time, far exceeds the physical in its extent and power. Indeed, it is the only one that matters. The world we live in is almost completely determined by it. You take it away, and you can still be alive. For millions of years human beings and their predecessors occupied the earth, but we would not want to live the lives that they had.

Esteemed reader. Please go back to Table 4.1 and take a few minutes to seriously study it. As said, all the items listed there are parts of a "pure" second reality. Not a single one exists in the physical universe. In your imagination strip them away, one by one, from the image you have of yourself, your "self," and see what is left. Much less than what you began with. Then multiply the loss by N times, N being an indeterminate but large number. The list in Table 4.1 is only a tiny sample of the total. Imagine what you have now. I suggest you have almost nothing.

If you feel you have succeeded in this task, you understand the story of the two realities.

5 MODEL FOR CONTROLLING FUNCTIONS

AAVO LUUK

This essay is devoted to an attempt to use the concept of function developed by Eric Hollnagel in his introduction of Functional Resonance Analysis Method (FRAM; Hollnagel, 2004; 2012) and his Extended Control Model (ECOM; Hollnagel, Nåbo, & Lau, 2003; Hollnagel, & Woods, 2005) to build a conceptual model to explain a potentially wide spectrum of psychological phenomena. Functions, with their goals and time-related control phases, create a set of means for a control-based approach of the modelling of the bio-psycho-social nature of the *Psyche* and its functioning in its diverse environments. The relevance of the topic to the present volume, dedicated to and in honor of my close friend and long-term colleague Jüri Allik, can be expressed through a question posed by Floyd Allport (1954): "What is latent or implicit *behind* molar formulations that can give the illumination which quantitative laws fail to provide?" (p. 284).

Introduction

The use of models in any domain of knowledge is essential for structuring the understanding and communication of the realm. As a rule, conceptual models represent topics of the world with the aim of using these in the development of systems of various scopes and kinds.

The present text has the aim of analyzing the sufficiently universal high-level concept of wide spectrum phenomena in psychology and beyond, called *functions* by Hollnagel (2012). The text has two main parts. In the first part, the characteristics of the function are explained and, in the second part, a model for controlling functions is presented. Explaining functions and their control through model-building may help create more or less universal means for understanding the time-dependent functioning of the bio-psycho-social entity of the *Psyche*.

Specifying Functions

Floyd Allport (1954, 1955), an older brother of one of the founding fathers of personality psychology, Gordon Allport, proposed a novel idea—modeling of psychological and other domains through events that can be used as dynamic structures for the analysis of psychological and other phenomena. Allport (1954, p. 288) explained:

E-mail: aavo.luuk@ut.ee

> *It is, rather, a dynamic structure—a structure of events ... Explanations must lie in the approximate "here and now" rather than in the remote past. The only way to accomplish this seems to be to cut across the conventional and absolute "time stream."... One can think of time as the duration occupied by the successive ongoing processes and events of a particular pattern that closes itself through a cycle of operation.*

The ideas of Allport (1954; 1955) have currency even today (see Morgeson, Mitchell, & Liu, 2015). Based on his experience of analyzing accidents and risk assessments in sociotechnical systems, Hollnagel (2004, 2012) reformulated the somewhat abstract dynamic event structure ideas of Allport into potentially universal units of analysis for a wide spectrum of psychological and other phenomena—the concept of function, which is equally suitable to the analysis of acting systems and people. Function according to Hollnagel (2012) seems to be a handier and more flexible and naturally applicable concept compared to Allport's event structures, considering the active, dynamic, developing nature of the processes which the two concepts have aimed to explain. Hollnagel (2012, p. 54) specifies the essence of functions in the following way:

> *Functions always describe something that can be done or is being done. The description is thus always of an activity, which in linguistic terms means that it must contain a verb phrase. The purpose of a function is to produce something or to bring about a state change. The result of a function, the Output, is always a description of a (system) state or of a condition.*

Hollnagel (2012) visualized his model of a function in the form of a regular-shaped hexagon with six features in its corners: Input, Output, Time, Control, Preconditions, and Resources. These features are considered sufficient for the further developing of interconnected structures—networks of functions—from separate functions by coupling them through features of the functions involved. Doing so, one can achieve networks of interrelated functions for searching for and following the potential risk-bearing couplings between these functions, which, in certain conditions, can lead to incidents or accidents. Both forward and backward tracing of risk-bearing couplings within the network for predicting future risks, or explaining preconditions of accidents that have already happened, is possible.

The present author has modified the visualization of the model of a single function, as shown in Figure 5.1. Modifications and developments in the understanding of the essence of a function are given in the three following sections.

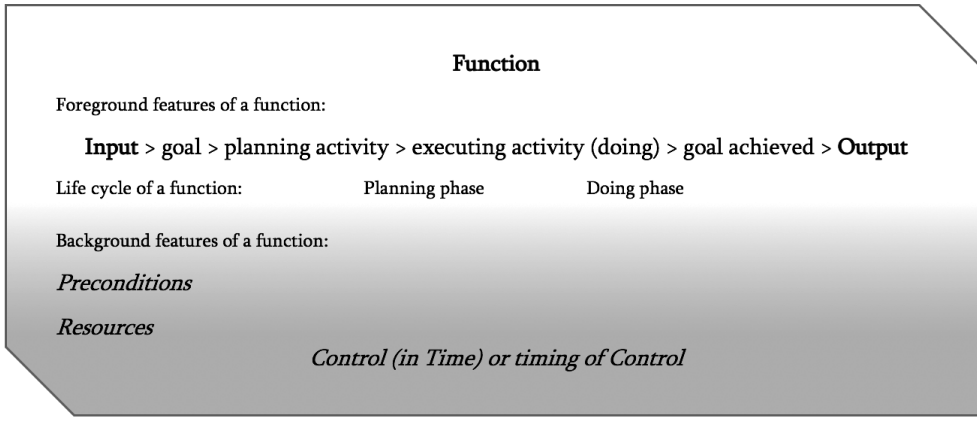


Figure 5.1 The author's modified graphical representation of the Hollnagel's (2012) model of a function.

United Time and Control

In the original model of Hollnagel (2012), Control and Time are considered separate features of a function for providing the necessary amount of potential and actual couplings with other functions, be they concurrent, past, or future, in relation to the function under consideration. This understanding can be characterized as an external view of the function, but if we take an internal view, then Control and Time together form a common background and backbone trait within the function that can be treated equally as control in Time or timing of Control. This combined trait can be considered the broadest background framing feature of the whole function, which exists, of course, together with the foreground activities of transforming Input into Output. Different to the internal view of Control and Time as separate features of a function, together with Preconditions and Resources, they remain appropriate for building the networks of coupled functions according to the external view of functions. There is no contradiction between internal and external views of the function. After introducing the function, Hollnagel (2012) has applied it solely in accordance with the external view of it, describing its relationships with other functions at systems and networks levels. The present author has not found any publications of Hollnagel, in which the constitution of the function has been analyzed from the internal point of view in the way it has been done in the present text.

Foreground and Background Features

Although Hollnagel (2012) writes with foresight about fore- and background features of functions elsewhere, in describing the applications of his Functional Resonance Analysis Method (FRAM), he does not distinguish between these two kinds of features within the networks of functions. It is reasonable to consider that features may emerge onto the foreground or recede into background, depending on the nature and lifecycle of the function and its couplings with other functions. Input and Output features of

functions should be mostly in the foreground during the transformation of the content of the former into the latter.

Differentiation of the foreground and background features of a function should not be understood as permanent and stable. Depending on the specific situation, initially typically background features (Time, Control, Preconditions, Resources) may selectively and temporarily shift into the foreground, and, due to this priority change inside the function, may result in the goal of the function as a whole becoming impeded, because the foreground activities of transforming the Input conditions into Output become blocked.

Planning and Doing Phases

The two potentially separable phases of planning and doing in a function are not directly presented in the model of Hollnagel (2012), but the need for doing so is repeatedly raised in the following citation (p. 38), as well as in his other books:

> *The distinction between work-as-imagined and work-as-done is often used in ergonomics literature to point out that there may be considerable difference between what people are assumed—or expected—to do and what they actually do. Work as imagined represents what designers, managers, regulators and authorities believe happens or should happen, whereas work-as-done represents what actually happens. Differences can either classified as noncompliance, violations, errors or as performance adjustments and improvisations, depending how one looks at it.*

According to the understanding of the present author, work-as-imagined may serve as Input to a function. As such, it may still need additional planning and adjustments at the start of the function, being transformed in the course of the function into work-as-done at the Output. Hollnagel (2012, p. 48) characterizes output as follows:

> *The Output from a function is the result of what the function does, for instance by processing the Input. ... The Output can be seen as representing a change of state—of the system or for one or more output parameters. ... The Output can, of course, also represent the signal that starts as a downstream function.*

A function, as a whole, may consist of planning and doing phases, or of only one of them. If there is no clear temporal separation between the two phases, or they are closely intertwined by nature, it remains to be decided if and when to consider the two as separate, or a single, united function. Preconditions typically change passively over time and actively under planning or executing activities, as do resources during the processing of the function. In the course of implementing a function, objective and subjective changes take place—the object of a function is transformed through active processes and a new understanding of the processes involved is developed in the actor. This means that two adjacent functions, even if similar, cannot simply be copies of each other, but contain important acquired differences.

Controlling Functions

The understanding of the function as developed by Hollnagel (2012) has good prospects to serve as a central and universal unit for potentially many conceptual models of the human *Psyche*. This prospect must be supported with the model of control of a function. Again, thanks to Hollnagel and colleagues (Hollnagel et al., 2003; Hollnagel, & Woods, 2005), an elaborate Extended Control Model (ECOM) is available and, after minimal adjustments, it can be applied to understand control inside and between the functions. The main difference between the ECOM developed by Hollnagel and colleagues (Hollnagel et al., 2003; Hollnagel, & Woods, 2005) and the understanding by the present author is in the scope of its application. In the present case, an attempt has been made to apply ECOM to explain control over the individual *Psyche*, predominantly at the level of its single functions, while for Hollnagel and colleagues (Hollnagel et al., 2003; Hollnagel, & Woods, 2005), the model covers any kind of system, but typically one with human participation. This particular difference, not to say limitation, brings us to another—in the control of systems, the stages or layers of control over system activities can change their order in time and also take place in a parallel manner, but in the individual *Psyche*, due to its cognitive and motor limitations, control layers can be understood almost exclusively as sequential features.

In his time, Allport (1954) was quite close to embedding the notion of control into his event structures, while writing about the self-closing or cyclical character of the structures of events. But, despite a clear understanding of the recurrent nature of event structures, the idea of control over them was not stated explicitly, and persists only implicitly in his understanding of the event-structure model.

From TOTE Model to ECOM

Miller, Galanter, and Pribram (1960) took control as the central issue in their approach by stating that the unit of analysis of behavior should be the feedback loop itself. Miller and colleagues (1960) believe that their Test-Operate-Test-Exit or TOTE unit, essentially incorporating feedback, can explain behavior in general (p. 32):

> *... the TOTE pattern describes both strategic and tactical units of behavior. Thus the operational phase of a higher-order TOTE might itself consist of a string of other TOTE units, and each of these, in turn, may contain still other strings of TOTEs, and so on.*

It seems that the pathos around the feedback principle raised in the book by Miller and colleagues (1960) and in other sources from those years has been quite influential and has inspired several generations of psychologists to search for and find feedback around almost all psychological matters.

The present author still considers the impact of Hollnagel's ideas (Hollnagel et al., 2003; Hollnagel & Woods, 2005) in applying the control principle in psychology more influential and far-reaching than the TOTE approach was or is. With their ECOM, Hollnagel and colleagues (2003; Hollnagel & Woods, 2005) equipped the explanation of the feedback principle with goals, phases, and timing of control. The graphic representation of his model (see Hollnagel, n.d.), reveals the recurrent, sequential and/or hierarchical nature of control, starting from the planning (targeting), and moving to the replanning (monitoring) and preparing execution (regulating), and, finally, to the executing (tracking) phases of any automatic or voluntary human activity (the original control phase terms developed by Hollnagel are given in parentheses). If, in artificial systems, control can be built in a hierarchical, parallel, sequential, or mixed manner (Lygeros, 1996), in individual human activities, the action to be controlled is accomplished predominantly sequentially, due to the limitations of human cognitive and motor abilities (Simon, 1955). Making the views of Hollnagel about functions applicable to building conceptual models of the individual *Psyche* requires their modification, by elaboration of the internal view of the function in addition to the external view of it, as developed in numerous works of Hollnagel. This is done here by including multi-level control and goal generation.

Function and its Control in Subjective Time

The best way to bind function with its control and time is to start with the life-cycle of the function. The following simplified attempt to visualize the controlling of a function in subjective time (Figure 5.2) may help one to grasp the idea. Depending on the nature of the function, at one extreme of the continuum of complexity, functions may be compressed in time, coherent, and simple to run, and it could take only seconds to complete them from scratch. At the other extreme of complexity, functions may be elongated in time, and include pauses or interruptions together with branching activities and resumptions on their way from Input to Output. The real duration of such a function may be immeasurable. Any function between these extremes has reasonable complexity and duration. Many, if not most, such functions have supposedly started in the Past, before the Present has arrived (Figure 5.2).

Faster by nature and coherent by structure, functions may fully fit into the Present, fill it, and even dictate its duration. At first glance, the idea of function determining the duration of the subjective Present may seem absurd, but if we think about the concept of the specious present, which was developed by Kelly (see Anonymous, 1882) and introduced to the psychological audience by James (1890), it may be worth considering. Many misinterpretations of this concept may arise from the fact that the elusive phenomenon of the specious (with the newer name subjective) present fades out in strict experimental tests of its nature and duration (White, 2017). But the disappearance of the specious present in experimentally planned, empirical testing conditions does not prove its nonexistence in more flexible everyday situations and

personal experiences. Subjectively, we do not feel that the present is over before we have completed the function at hand. The duration of the subjective present depends on the duration of the time we spend on the actual, presently active function on which we are working. The same should hold true for the different control phases in running the function: if its coherence is low, the phases of control are separated in time and each of them requires considerable duration to complete.

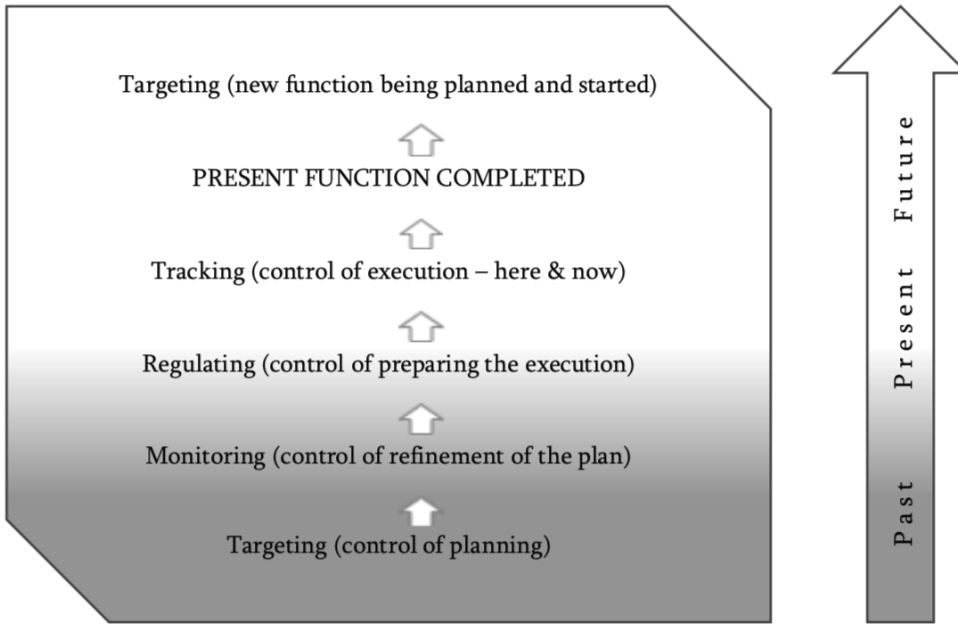


Figure 5.2 An example of the control of a function in subjective time.

In the individual psychological context, a function binds three categories of subjective time from Past through Present to the Future. It is clear that an actual function can be handled during the Present only. Together with the idea that the duration of the function determines the duration of the subjective Present, it is also evident that labor-intensive and long-term functions may have been started in the Past, with certain segments of them being worked out during the Present, and the completion of the function still to be handed over to the Future present.

Mode of Control

One important corollary of using a subjective timeline of Past, Present, and Future in the control of a function is related to the mode of control. From control theory, it is evident that control can typically be achieved by using closed-loop feedback, open-loop control, or their combination, termed feedforward. It may be helpful to consider

the feedback mode of control as an online process, while open-loop and feedforward modes of control could be characterized as offline processes. In online control, after detecting a change (i.e., variability), the unit responsible for control will react—act retroactively—afterwards, while, in open loop and feedforward modes of control, the unit in control has been adjusted to influence the processes proactively—prospectively (often long before the urgent need to influence the processes may appear), typically through predefined—prescribed—limits of variability. Different theoretical and applied branches of control theory started with modeling and application of the feedback principle, while efforts to understand feedforward are comparatively new. In reality, feedback and feedforward are two interrelated aspects of closed-loop control and one cannot exist without the other.

Due to its complexity (and uncertainty), feedforward initially received less attention in control theory than feedback, but, in human relations and activities on a large scale, the importance of feedforward should be considered greater than that of a feedback. If we try to rely on Figure 5.2 for understanding the essence of feedforward, we must admit that potentially everything relevant from our bio-psycho-social past (experiences of any kind from any domain together with our innate dispositions), from the acute environment, and from previous functions, form the grounds for initiating (targeting) a function. This "potentially everything" has to be analyzed and categorized for each concrete function to identify the most relevant preconditions and resources for this function. Through recurrent applying of cognitive, motor, and affective processes in the feedback mode, goals and means for achieving them are generated (targeting), elaborated on (monitoring), repeatedly and with increasing precision prepared for execution (regulating), and finally executed (tracking) to complete the function (here, again, the original control phase terms, developed by Hollnagel, 2012, are in parentheses). The recurrent nature of generating endless numbers of new functions within repeated control phases should also be noted. The feedforward mode of control provides valuable initial raw material for a function and the feedback mode of control carries out the mundane duties of processing this material into the highly valuable and sophisticated end-product at the Output of a function.

Generation of Routine Functions

Discussion of the control over functions would be too limited if we did not consider the ways in which functions are generated. The emerging of functions is inextricable from the origination of their goals. For clarity, we consider only two kinds of functions—feature-dependent or routine or automatic, and voluntarily created, functions.

Feature-dependent are the functions that own their well-established place in the network of related functions due to their essential use of schemas and scripts. Feature-dependence originates from dependence on the preconditions, resources, timing, and control of the present function. On some occasions, a feature-dependent function may

give birth to several functions coupled to the present or sub- or branching functions. Such feature-dependent functions could also be called schema- and script-based or routine functions. Most probably, such functions serve the everyday and routine needs and habits of our bio-psycho-social existence, without the immediate need to involve extensive or intensive cognitive, motor, or emotional efforts devoted to the planning phase of this function. Put another way, the history of generating goals of routine functions lies somewhere far in the past and, since then, these goals have been used and reused an endless number of times, becoming, finally, automatisms. The idea of automatisms can be equally applied to functions and their goals and means. When goals and means for achieving them are known and available beforehand, there is little need for planning, and the main efforts should be devoted to doing the activities of the function.

If we take the cooking of an everyday meal as an example of this kind of function, it is quite easy to retrieve from memory or from a cookbook the lists of requirements for a successful start, running, and completing of a specific meal-cooking function. And it is equally easy to imagine the growth of subfunctions or branching functions, if something in reality does not fit well into preexisting schemas and scripts of cooking this specific kind of meal.

Outliers from schemas and scripts among resources, preconditions, timing, or control of a routine function convert the till-now routine function into a new voluntary function. These outliers will, at least, be related to information, space, time, and energy.

Initiation of Voluntary Functions

Differently to routine functions, where the goal to be achieved does not demand special efforts for its generation and does not reside in the foreground of the function most of time, voluntarily created functions have foreground goals and means for fulfilling them developed over the course of resource-demanding motivational efforts.

A very brief excursion into the realm of motivational psychology may be beneficial for better understanding goals. Klinger (1975, 1977; Klinger & Cox, 2004) was among the first influential psychologists in the pithy introduction of the concept of goals into the psychology of motivation. Incentives—the events and objects valued—which the individual commits to pursuing, become goals. Numerous goals (also called current concerns by Klinger) may exist simultaneously, but only one is actively pursued at a time.

Recently, considerable progress in understanding goals and means for achieving them has been achieved by Kruglanski and colleagues (2002), who see motivation as a cognition, at the center of which are goal systems "as the mental representations of motivational networks composed of interconnected goals and means" (p. 333). Goal systems are particularly useful when considering multiple goals and their relationships together with their means. An additional impact of Kruglanski et al. (2015) is their

explanation of the multistep sequence from an attitude to a goal and then on to a behavior to achieve this goal.

According to Kruglanski and colleagues (2015), attitude or relative liking generates wanting or desirability, which is conjoined with expected attainability. Desirability/attainability of a sufficient magnitude generates goal commitment. To initiate a behavior, the goal must currently be active and dominant over alternatives, and the behavior chosen must be the preferred means to the goal. While Klinger (1975, 1977) is content with the explanation how commitment to incentive produces a goal, Kruglanski and colleagues (2015) look into the full chain of steps from attitude to goal and further to behavioral means for achieving the goal. As he explains, the sequence of events in the chain can serve as a model input into the generation of goals for voluntary functions.

Conclusion

In accordance with the conceptual model of the control of functions developed in the present essay, goals guide present and immediate future actions, both between and within functions, keeping our present bio-psycho-social activities under way, and this has been laid down by previously completed functions and is continued by successive phases of the present one. Goals serve the important duty of uniting and cumulating our efforts at different time segments into a whole sequence of persistent goal-striving endeavors, through our personal past, present, and future. Functions, with their goals and time-related control phases, create a set of means for control-based attempts at modelling the functioning of the *Psyche*.

References

Allport, F. H. (1954). The structuring of events: Outline of a general theory with applications to psychology. *Psychological Review, 61*, 281–303.

Allport, F. H. (1955). *Theories of perception and the concept of structure.* New York: John Wiley and Sons, Inc.

Anonymous [E. Robert Kelly, "E. R. Clay"] (1882). *The alternative: A study in psychology*. London: Macmillan and Co.

Hollnagel, E. (2004). *Barriers and accident prevention.* Aldershot: Ashgate Publishing Limited.

Hollnagel, E. (2012). *FRAM: The functional resonance analysis method. Modelling complex socio-technical systems*. Farnham, UK: Ashgate.

Hollnagel, E. (n.d.). *Extended Control Model (ECOM).* Retrieved December 6, 2018 from http://erikhollnagel.com/onewebmedia/ECOM.pdf

Hollnagel, E., Nåbo, A. & Lau, I. V. (2003). A systematic model of driver-in-control. In J. D. Lee, M. Rizzo, and D. V. McGehee (Eds.), *Proceedings of the second international driving symposium on human factors in driver assessment, training and vehicle design* (pp. 86–91). Iowa City, UT: University of Iowa.

Hollnagel, E. & Woods, D. D. (2005). *Joint cognitive systems: Foundations of cognitive systems engineering.* Boca Raton, FL: CRC Press/Taylor & Francis.

James, W. (1890). *The principles of psychology.* New York: Henry Holt & Company.

Klinger, E. (1975). Consequences of commitment to and disengagement from incentives. *Psychological Review, 82*, 1–25.

Klinger, E. (1977). *Meaning and void: Inner experience and the incentives in people's lives*. Minneapolis, MN: University of Minnesota Press.

Klinger, E., & Cox, W. M. (2004). Motivation and the theory of current concerns. In W. M. Cox, & E. Klinger (Eds.), *Handbook of motivational counseling: Concepts, approaches and assessment* (pp. 3–27). Chichester: John Wiley & Sons.

Kruglanski, A. W., Shah, J. Y., Fishbach, A., Friedman, R., Chun, W. Y., & Sleeth-Keppler, D. (2002). A theory of goal systems. *Advances in Experimental Social Psychology, 34*, 331–377.

Kruglanski, A. W., Jasko, K., Chernikova, M., Milyavsky, M., Babush, M., Baldner, C., & Pierro, A. (2015). The rocky road from attitudes to behaviors: Charting the goal systemic course of actions. *Psychological Review, 122*, 598–620.

Lygeros, J. (1996). *Hierarchical, hybrid control of large scale systems.* (California PATH Research Report UCB-ITS-PRR-96–23). Berkeley, CA: University of California.

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior.* New York: Holt, Rinehart & Winston, Inc.

Morgeson, F. P., Mitchell, T. R., & Liu, D. (2015). Event system theory: An event-oriented approach to the organizational sciences. *Academy of Management Review, 40*, 515–537.

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics, 69*, 99–118.

White, P. A. (2017). The three-second "subjective present": A critical review and a new proposal. *Psychological Bulletin, 143*, 735–756.

6 FOOD FOR THOUGHT

ROBERT R. MCCRAE
Gloucester, Massachusetts, USA

Five-Factor Theory (FFT) is a general theory of personality developed to account for research findings on the Five-Factor Model of personality traits, including its universality. The distinction between biologically-based Basic Tendencies (BTs) and acquired Characteristic Adaptations (CAs) is central to FFT. In this chapter I elaborate on that distinction and address the issue of duality: Personality scales are simultaneously direct measures of CAs and indirect indicators of BTs. Implicit measures of traits, although potentially useful, do not circumvent duality, and FFT presumes that people have no direct intuition of their traits. Finally, I discuss FFT's powerful and parsimonious but unpopular assumption that traits are not altered by the psychological environment.

Introduction

Since my retirement, I have had more time to think about one of my favorite topics—food—and have noticed something remarkable: The grand cuisines of France and China, the simple diets of the Inuit and Yanomami, the typical fare at a New York deli or a Texas barbecue are all essentially the same. Everyone eats the same foods, and these foods fall into categories, of which there happen to be . . . five: vitamins, minerals, proteins, fats, and carbohydrates. This is the human diet (incidentally, it is also more or less the mammalian diet). However, it is equally remarkable that every culture finds its own unique ways to serve up these nutrients—ways that reflect the local flora and fauna, religious taboos, cooking techniques, and so on. One cannot hope to understand *food* without considering both these levels: its universal function as human nutrition, and its deliciously diverse manifestations in different cuisines.

I will argue that there is a striking parallel in psychology. Features of human personality can also be construed on two levels, *Basic Tendencies* (BTs) and *Characteristic Adaptations* (CAs). The former are a part of universal human nature; the latter are specific to time and place. These are the central concepts in the Five-Factor Theory (FFT; McCrae & Costa, 2008) of personality. FFT is widely cited, but—with a few notable exceptions (e.g., Allik, 2002)—it has not been widely embraced. There

Portions of this chapter were presented as the Jack Block Award Address, 15th Annual Meeting of the Society for Personality and Social Psychology, February 2014, Austin, TX. I thank Gina Sutin, Claudio Barbaranelli, and Colin DeYoung for their assistance. Robert R. McCrae receives royalties from the NEO Inventories.

E-mail: rrmccrae@gmail.com

seem to be two features of the theory that pose difficulties for some psychologists: the conceptual distinction between BTs and CAs; and the postulate that BTs are not affected by the psychological environment. Here I address those issues.

Five-Factor Theory

Anthropologists like Ruth Benedict (1934), who were the first to study personality and culture, tended to argue that human nature was plastic, and personality was created through the process of enculturation. Because cultures show a wide range of profound differences, most social scientists assumed that the personality psychology of Americans would not be generalizable to other languages and cultures; instead, a host of distinct indigenous psychologies would be needed. That is, of course, a testable hypothesis, and thanks to the efforts of scores of translators of personality trait measures (e.g., Kallasmaa, Allik, Realo, & McCrae, 2000), it has now been tested in countries around the world. The hypothesis was not supported; to the contrary, a large and consistent body of data has clearly demonstrated that traits, and many of their psychometric characteristics, are universal (Allik, Realo, & McCrae, 2013; Allik et al., 2010). It appears that personality traits are not written by culture on a tabula rasa; they are instead an intrinsic part of human nature, built into our genes and our brains.

That, in any case, is the interpretation offered by FFT (McCrae & Costa, 2008), a general personality theory proposed to account for the body of findings that research on the Five-Factor Model has generated. But the theory must explain more than universals; it must also account for the fact that personality is expressed quite differently in different cultures. For example, an extraverted American woman is likely to have many opposite-sex friends, whereas an extraverted woman from a fundamentalist Islamic culture would not be allowed to; her warmth and gregariousness would need to be expressed within her family or among her female friends. Again, people high in Neuroticism tend to worry, but whether they worry about insider trading or plagues of locusts depends on their life circumstances. In FFT, we say that traits are BTs, whereas the beliefs, attitudes, habits, and relationships that express them are CAs.

Figure 6.1 shows a schematic version of FFT; the central components of the personality system are BTs and CAs. CAs, at least, are familiar to psychologists: They include skills, attitudes, behavioral routines, roles and relationships, and the self-concept. They are called *adaptations* because they are acquired as the individual interacts with, and adapts to, the opportunities and requirements of the social environment. They are *characteristic* insofar as they are also shaped by, and thus express, the distinctive personality traits of the individual. For example, all students have to do homework, but the study habits of conscientious students are characteristically different from those developed by their less highly motivated peers.

Although it may seem odd to fit such diverse phenomena as relationships, motor skills, and vocational interests into a single category, at least the contents are readily understandable.

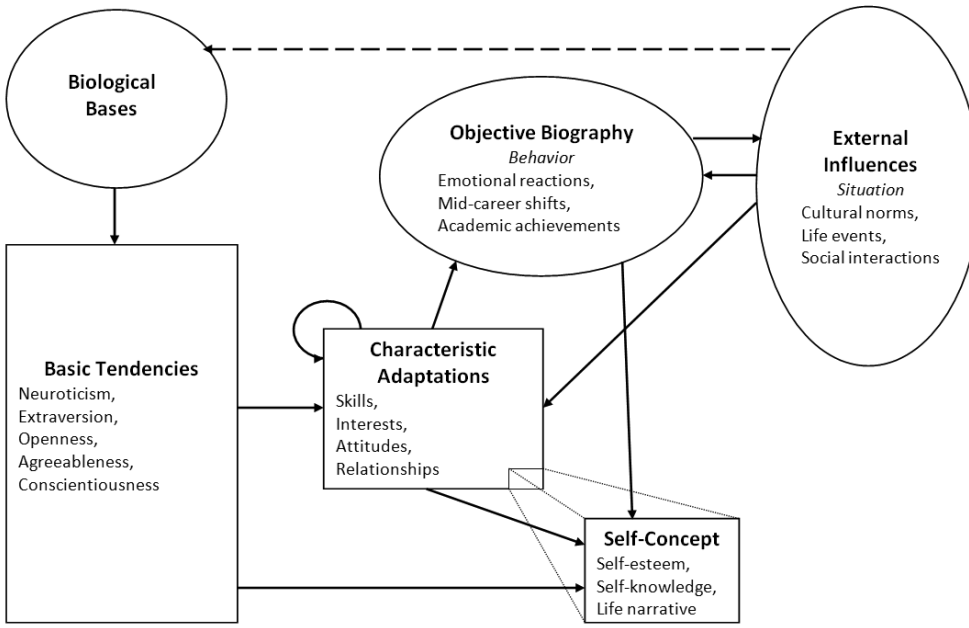


Figure 6.1 A representation of the Five-Factor Theory (FFT) personality system. Core components are in rectangles; interfacing components are in ellipses; arrows represent causal pathways on which dynamic processes operate. Adapted from McCrae and Costa (2008).

This is not necessarily so for BTs, especially personality traits. Many psychologists are accustomed to thinking of traits as overt patterns of behavior. For them, to say that people are high in the trait of Order is to say that they typically keep their desk neat, promptly write their engagements on a calendar, and shine their shoes daily. From this perspective, traits are surely effects rather than causes, and attempts to explain behavior from trait levels is misguided (cf. Cervone, 2005). From this perspective, also, traits are culture-specific: It is unlikely that the Yanomami keep neat desks, up-to-date calendars, or polished shoes, so how could they show the trait of Order? Clearly, FFT has a completely different conception of traits.

DeYoung (2015) comes a bit closer with his definition of traits when he "equates traits with the tendency to be in certain emotional, motivational, cognitive, and behavioral states" (p. 35). The key word here is "tendency," which may suggest an underlying causal structure that generates the states. But DeYoung (2015) also says that traits are "probabilistic descriptions of relatively stable patterns of emotion, motivation,

cognition, and behavior" (p. 35, italics omitted), which sounds more descriptive than explanatory. FFT insists upon a different conception of traits: They are unobserved psychological constructs that correspond to a causal basis of observed patterns of emotion, motivation, cognition, and behavior.

Psychologists are familiar with this kind of construct from cognitive psychology. Intelligence (which FFT considers a BT) is surely not a pattern of right and wrong answers on an IQ test; it is a hypothetical ability that, with a given educational history, allows an individual to make some number of correct answers.

However, when we assess intelligence, we often do it through an achievement test, such as a vocabulary test. Knowing the meaning of words is not the same as being smart, but most people with large vocabularies in fact have high IQs, because intelligence facilitates word learning. Similarly, when we assess personality, we often infer traits by asking people about their CAs. For example, if we want to know if someone is high or low on Gregariousness, we can ask them about the kind of work preferences they have (e.g., salesperson vs. forest ranger), because these preferences are adaptations shaped in part by the characteristic level of Gregariousness. But just as intelligence is distinct from the achievement we use to measure it, traits are categorically distinct from the CAs that express them—in the same way that vitamins and proteins are categorically different kinds of food than Caesar salad and Eggs Benedict. BTs are abstract potentials that are indicated by their concrete manifestations acquired in a specific social environment.

The Duality Principle

The notion that traits can be inferred from their outward expressions in actions, values, motives, and so on is not in itself problematic. It is a simple instance of the inductive reasoning that laypersons and scientists alike rely on constantly: "Extraverts attend parties; X attends parties; therefore, X is (probably) an extravert." The logician's distaste for induction stems from that telltale "probably;" there are, of course, other possible reasons why an individual would attend parties. Perhaps X is an introvert who desperately needs a job and finds social networking essential. Perhaps X is an author doing research for a book on Extraversion. Perhaps X is dragged to parties by a spouse.

The personality questionnaire item, "I often go to parties," is therefore susceptible of two interpretations: It may simply be a 'probabilistic description of a relatively stable pattern of behavior' which might have any number of causes; or it may be an indicator of Extraversion. We refer to this as the duality principle (Costa & McCrae, 2017), and we have argued elsewhere (McCrae & Sutin, 2018) that this is always the case for personality items, even those which seem to be direct assessments of a trait, such as "I am an extravert."

Psychologists have learned to live with imperfect assessments of constructs, but they prefer to check them against alternative methods of measurement. Here duality poses a problem. We cannot check self-reports by consulting informant ratings, because observers witness the same ambiguous behavior: Yes, X has been observed going to parties, but what does it mean?

As Figure 6.1 shows, there are only two pathways to BTs: From biological bases, and to CAs (including the Self-Concept). CAs are ambiguous as trait indicators, because there is another pathway from External Influences to CAs, and that may be the causally decisive factor. It seems the only remaining possibility is to assess the biological basis of traits. In principle this is an appropriate method; in practice, we are very far from understanding either the genetic bases of traits or their more proximal mechanisms in brain structure and functioning.

Implicit Measures

If questionnaire items are ambiguous, what about implicit trait measures (Dentale, Vecchione, & Barbaranelli, 2016)? Assessing traits with the Implicit Association Test (IAT) is supposed to offer an alternative to the usual conscious self-reports; because they rely on the automatic speed of cognitive processing, implicit measures may be less susceptible to distortion. Research on the assessment of FFM traits using the IAT is at least mildly encouraging: IAT measures of the five domains show convergent and discriminant validity when correlated with explicit, self-report questionnaire measures, although the magnitude of the convergent correlations is small ($r \approx .20$; De Cuyper et al., 2017).[1] IAT measures sometimes, but not always, predict the same behavioral criteria as their explicit counterparts (Back, Schmukle, & Egloff, 2009). Internal consistency is high ($\alpha \approx .80$), although retest reliability is low ($r_{tt} \approx .50$; Teige-Mocigemba, Klauer, & Sherman, 2010), suggesting that IATs are in part state measures.

In the IAT paradigm, respondents are presented with pairs of stimuli and asked to assign them to one of two categories using keys on the left or right of the keyboard. For example, given the pair of stimuli [me] and [party-going], they may be asked to assign it to the left category which is defined as "Self + Extraverted." In theory, if the respondent is extraverted, this will be an easy decision, because [me] and [party-going] seems a natural combination. It will take longer for an introvert to make the decision, because the combination seems odd and confusing. A metric based on decision times is thus used to infer a trait.

IATs are promising for certain applications (Vecchione, Dentale, Alessandri & Barbaranelli, 2014), but they probably do not solve the duality problem. To an introvert who has been coerced into attending a large number of parties, [me] and [party-going] may be a sadly familiar combination that elicits a rapid response. Both implicit and

[1] Curiously, no one seems to have validated IAT measures against informant ratings of traits.

explicit assessments speak directly to the CA, and only indirectly and probabilistically to the intended trait.[2]

Introspective Intuition

Individuals have access to an entire category of data that are unavailable to external observers: internal thoughts, feelings, and wishes. It is tempting to think that these are direct representations of personality traits, and thus that individuals have an intuitive grasp of their real personality. This is not the interpretation that FFT offers. From the perspective of FFT, private thoughts, feelings, and wishes are simply part of the Objective Biography. The individual (though not outside observers) can draw on these private experiences to make inferences about CAs and ultimately about underlying traits, but these inferences are in principle no different from those drawn from patterns of overt speech and behavior.

Intuitive self-knowledge may be suggested to some readers by the arrow in Figure 6.1 leading directly from BTs to the Self-Concept. If so, a clarification is needed, because that is not what FFT proposes—that is not it, at all. The arrow is intended to show that traits influence the content of the Self-Concept, which is a highly selective account of the individual's characteristics; the selections are determined in part by traits. Individuals high in Neuroticism may emphasize their failings, whereas those low in Modesty will highlight their perceived superiority. The Self-Concept is a CA, and, like other CAs, it is shaped by traits. But it is not a direct representation of those traits.

FFT holds that self-knowledge, like all knowledge, must be based on experience. Suppose an individual who has never encountered the fruit were asked if the smell of durian (https://en.wikipedia.org/wiki/Duriana) is pleasantly sweet or disgustingly vile. One of these is almost certain to be true, and given a whiff, the answer is immediately obvious. But people cannot grasp the answer by sheer intuition of their innate sense of smell in the absence of experiential input. The same is true of traits.

Aggregation

So, what is to be done? As so often in psychometrics, the answer lies in aggregation—not over raters, but over trait indicators. Any given indicator, such as party-going, may be ambiguous, but when many different indicators of the trait are assessed and averaged, the environmental accidents that help shape behaviors tend to cancel out. How likely is it that someone who attends parties only to research a book also rides roller coasters only to impress a potential mate, and laughs loudly only because of

[2] Implicit measures sometime show incremental validity beyond self-reports in the prediction of behaviors, and this is sometimes interpreted to mean that they "capture unique information that was not provided by a self-report scale" (Dentale, Vecchione, & Barbaranelli, 2016, p. 120). From the perspective of FFT, such an interpretation implies that implicit traits are distinct facets of the trait domains; they might, for example, be uniquely related to spontaneous behaviors (Asendorpf, Banse, & Mücke, 2002). A simpler view is that, like informant ratings, implicit measures assess the same trait as explicit measures, but differ in method biases and error (McCrae, 2018), and thus contribute through aggregation of the true score.

impaired hearing? It is easier to believe that a gregarious, thrill-seeking, boisterous person really is an extravert.

Causal Direction

Five Factor Theory goes beyond conceptual distinctions to posit a particular causal structure among the components of the personality system. In Figure 6.1, the allowed causal pathways are indicated by arrows.

CAs are influenced by both BTs and the social environment. That is not controversial. Nor it is revolutionary to say that BTs have some biological basis, as the enormous investment in the search for personality genes attests. What sets FFT apart from most other theories of personality is that it denies any influence of the psychological environment on personality traits: There is no arrow from External Influences directly to BTs.

This is the most controversial aspect of Five Factor Theory, and it is surely an oversimplification. However, I think it is a remarkably powerful approximation to the truth (McCrae, De Bolle, Löckenhoff, & Terracciano, in press). If personality traits are really insulated from environmental influences, they are likely to appear and persist in the same form despite a host of variations in the circumstances in which they are encountered, and several different lines of evidence suggest that this is just what they do. This postulate of FFT explains why personality changes so little across the decades of adulthood, despite stressful life events, role changes, and years of watching television (McCrae & Costa, 2003): Such experiences apparently have little lasting influence on traits. It explains why behavior genetic studies find so little evidence of shared environmental effects (Plomin & Daniels, 1987). It explains why historical events like wars and social upheaval leave so little imprint on the personality profiles of different age cohorts (Costa & McCrae, 2000). And of course, it explains why the Five-Factor Model is a good description of personality in the diverse cultures of the world (McCrae, Terracciano, & 78 Members of the Personality Profiles of Cultures Project, 2005).

An adept at path analysis might object that Figure 6.1 allows one pathway from External Influences to BTs, via the dashed arrow into Biological Bases. That arrow was added some years after the original model to acknowledge that external forces can affect traits when they alter the brain itself. Malnutrition (Galler et al., 2013), disease (Siegler, Dawson, & Welsh, 1994), and physical activity (Stephan, Sutin, & Terracciano, 2014), as well as psychopharmacological interventions (Costa, Bagby, Herbst, & McCrae, 2005), provide examples.

DeYoung (2015) argued that all "psychological processes supervene on biological processes" (p. 33)—a more sophisticated statement of Murray's adage, "no brain, no personality." One implication of this premise is that it would be meaningful to add additional pathways to Figure 6.1 showing that the effects of External Influences

on CAs and on behaviors are mediated by Biological Bases. A red light evokes our knowledge of traffic rules and our habit of stopping, but it does so only because it is sensed by the eye, interpreted by the visual cortex, relayed to the relevant motor area, and so on. These pathways are not included in the figure because they are (from FFT's perspective) trivially true.

DeYoung (2015), however, proposed another, and certainly not trivial, implication. He suggested that the dashed arrow in Figure 6.1 opens the floodgates to environmental influences on traits, because "many experiences cause analogous neurobiological changes that could lead to lasting changes in traits" (p. 38, Footnote 5). Yes, experiences do cause changes in the brain; all acquired CAs presumably are encoded neurobiologically. But whether these changes are truly "analogous" to those induced by disease or drugs—whether they can lead to changes in underlying personality traits—is arguable (Costa, McCrae, & Löckenhoff, 2019). FFT assumes that CAs are like software, which can be programmed, whereas BTs are like hardware, which can be changed only by a physical modification of the equipment. Perhaps the most crucial test of this premise of FFT is psychotherapy, which may (Roberts et al., 2017) or may not (Chow, Wagner, Lüdtke, Trautwein, & Roberts, 2017) modify trait levels. If it does modify traits, it constitutes a challenge to FFT that will eventually require modifications to the theory.

In the meantime, the view that traits are immune to the effects of the psychological environment provides a powerful and parsimonious explanation of a host of well-replicated observations. It is not popular, however, for two reasons. First, almost all prior theories of personality have assumed that it is the psychological environment that shapes us—whether though traumas of early childhood, or histories of reinforcement, or cultural-historical imperatives. The idea that personality traits would be beyond the reach of such powerful influences runs counter to everything psychologists have been taught. One of the reasons FFT insists on purely biological influences on traits is strategic: It attempts to dislodge this default environmental assumption so that psychologists can more impartially investigate the issue.

Second, many psychologists are unhappy with FFT because it suggests limits to psychological interventions. Traits are important determinants of a wide range of life outcomes, so it would be wonderful to be able to manipulate them. If we could reduce Neuroticism, we could cure personality disorders; if we could increase Openness, we could eliminate prejudice; if we could instill Conscientiousness, we could enhance learning in school and productivity at work. But, per FFT, we can't. Instead, we must be content to reduce distress long enough to avert a crisis; foster realistic expectations for change; channel the expression of undesirable traits in more socially acceptable directions; and match individuals to tasks that are consistent with their personality profiles. These are worthy and achievable goals and should be the focus of therapists and applied psychologists. But they lack the glamour of the wholesale conversion of personality that modifying traits seems to promise.

For many readers, I fear that I will have ended this essay with a metaphorical dose of cod liver oil—salubrious but distasteful—instead of the usual crème brûlée. But the same psychological organization that makes individuals resistant to the attempted interventions of psychologists also gives them the ability to maintain their identity in the face of childhood deprivations, peer pressure, and totalitarian regimes. Traits form an important component of one's identity (McCrae & Costa, 1988), and, like it or not, they are here to stay.

References

Allik, J. (2002). Towards a theory of personality. In C. von Hofsten & L. Bäckman (Eds.), *Psychology at the turn of the millennium* (Vol. 2, pp. 179–200). New York: Taylor & Francis.

Allik, J., Realo, A., & McCrae, R. R. (2013). Universality of the Five-Factor Model of personality. In T. A. Widiger & P. T. Costa, Jr. (Eds.), *Personality disorders and the Five-Factor Model of personality* (3rd ed., pp. 61–74). Washington, DC: American Psychological Association.

Allik, J., Realo, A., Mõttus, R., Borkenau, P., Kuppens, P., & Hřebíčková, M. (2010). How people see others is different from how people see themselves: A replicable pattern across cultures. *Journal of Personality and Social Psychology, 99*, 870–882.

Asendorpf, J. B., Banse, R., & Mücke, D. (2002). Double dissociation between implicit and explicit personality self-concept: The case of shy behavior. *Journal of Personality and Social Psychology, 83*, 380–393.

Back, M. D., Schmukle, S. C., & Egloff, B. (2009). Predicting actual behavior from the explicit and implicit self-concept of personality. *Journal of Personality and Social Psychology, 97,* 533–548.

Benedict, R. (1934). *Patterns of culture*. Boston: Houghton-Mifflin.

Cervone, D. (2005). Personality architecture: Within-person structures and processes. *Annual Review of Psychology, 56*, 423–452.

Chow, P. I., Wagner, J., Lüdtke, O., Trautwein, U., & Roberts, B. W. (2017). Therapy experience in naturalistic observational studies is associated with negative changes in personality. *Journal of Research in Personality*, *68*, 88–95.

Costa, P. T., Jr., Bagby, R. M., Herbst, J. H., & McCrae, R. R. (2005). Personality self-reports are concurrently reliable and valid during acute depressive episodes. *Journal of Affective Disorders*, *89*, 45–55.

Costa, P. T., Jr., & McCrae, R. R. (2000). A theoretical context for adult temperament. In T. D. Wachs & G. A. Kohnstamm (Eds.), *Temperament in context* (pp. 1–21). Hillsdale, NJ: Lawrence Erlbaum Associates.

Costa, P. T., Jr., & McCrae, R. R. (2017). The NEO Inventories as instruments of psychological theory. In T. A. Widiger (Ed.), *Oxford handbook of the Five-Factor Model* (pp. 11–37). New York: Oxford University Press.

Costa, P. T., Jr., McCrae, R. R., & Löckenhoff, C. E. (2019). Personality across the life span. *Annual Review of Psychology*, *70*, 423–448.

De Cuyper, K., De Houwer, J., Vansteelandt, K., Perugini, M., Pieters, G., Claes, L., & Hermans, D. (2017). Using indirect measurement tasks to assess the self-concept of personality: A systematic review and meta-analyses. *European Journal of Personality, 31*, 8–41.

Dentale, F., Vecchione, M., & Barbaranelli, C. (2016). Applying the IAT to assess Big Five personality traits: A brief review of measurement and validity issues. In Z. Jin (Ed.), *Psychology and mental health: Concepts, methodologies, tools, and applications* (Vol. 1., pp. 113–127). Hershey, PA: IGI Global.

DeYoung, C. (2015). Cybernetic Big Five Theory. *Journal of Research in Personality, 56*, 33–58.

Galler, J. R., Bryce, C. P., Zichlin, M. L., Waber, D. P., Exner, N., Fitzmaurice, G. M., & Costa, P. T., Jr. (2013). Malnutrition in the first year of life and personality at age 40. *Journal of Child Psychology and Psychiatry*, *54*, 911–918.

Kallasmaa, T., Allik, J., Realo, A., & McCrae, R. R. (2000). The Estonian version of the NEO PI-R: An examination of universal and culture-specific aspects of the Five-Factor Model. *European Journal of Personality*, *14*, 265–278.

McCrae, R. R. (2018). Method biases in single-source personality assessments. *Psychological Assessment, 30,* 1160–1173.

McCrae, R. R., & Costa, P. T., Jr. (1988). Age, personality, and the spontaneous self-concept. *Journal of Gerontology: Social Sciences*, *43B*, S177–S185.

McCrae, R. R., & Costa, P. T., Jr. (2003). *Personality in adulthood: A Five-Factor Theory perspective* (2nd ed.). New York: Guilford.

McCrae, R. R., & Costa, P. T., Jr. (2008). The Five-Factor Theory of personality. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 159–181). New York: Guilford.

McCrae, R. R., De Bolle, M., Löckenhoff, C. E., & Terracciano, A. (in press). Lifespan trait development: Towards an adequate theory of personality. In J. F. Rauthmann (Ed.), *Handbook of personality dynamics and processes*. Amsterdam: Elsevier.

McCrae, R. R., & Sutin, A. R. (2018). A Five-Factor Theory perspective on causal analysis. *European Journal of Personality*, *32*, 151–166.

McCrae, R. R., Terracciano, A., & 78 Members of the Personality Profiles of Cultures Project. (2005). Universal features of personality traits from the observer's perspective: Data from 50 cultures. *Journal of Personality and Social Psychology, 88,* 547–561

Plomin, R., & Daniels, D. (1987). Why are children in the same family so different from one another? *Behavioral and Brain Sciences, 10*, 1–16.

Roberts, B. W., Luo, J., Briley, D. A., Chow, P. I., Su, R., & Hill, P. L. (2017). A systematic review of personality trait change through intervention. *Psychological Bulletin*, *143*, 117–141.

Siegler, I. C., Dawson, D. V., & Welsh, K. A. (1994). Caregiver ratings of personality change in Alzheimer's disease patients: A replication. *Psychology and Aging, 9,* 464–466.

Stephan, Y., Sutin, A. R., & Terracciano, A. (2014). Physical activity and personality development across adulthood and old age: Evidence from two longitudinal studies. *Journal of Research in Personality*, *49*, 1–7.

Teige-Mocigemba, S., Klauer, K. C., & Sherman, J. W. (2010). Practical guide to Implicit Association Test and related tasks. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 117–139). New York: Guilford Press.

Vecchione, M., Dentale, F., Alessandri, G. & Barbaranelli, C. (2014). Fakeability of implicit and explicit measures of the Big Five: Research findings from organizational settings. *International Journal of Selection and Assessment, 22*, 211–218.

7 SOCIAL DESIRABILITY FROM A PRAGMATIC POINT OF VIEW

KENN KONSTABEL

Social desirability in personality inventories is usually treated as (a) a question of the connotative meaning (evaluativeness) of items, and/or (b) a self-enhancing (evaluative) bias in responses, and (c) possibly an interactive effect of the two. Using Jakobson's (1960) classification of the functions of language, I argue that, in addition to the above possibilities, social desirability (that is, agreeing with desirable items and disagreeing with undesirable ones) can stem from at least four distinct functions of language (referential, emotive, conative, and phatic). A socially desirable response may, for instance, reflect the desirable qualities of the target, the respondent's understanding that the desirable qualities are relevant to the purpose of communication (referential function), the respondent's motivated beliefs about the target (emotive function), the respondent's goal to influence the test user's beliefs in a desirable direction (conative function), or be a normative, easy-to-give answer (phatic function). Further, a socially desirable response can serve autocommunicative purposes, reminding the respondent of a socially desirable belief or action that they would like to have or perform.

Pragmatics

In common use, to say that something is pragmatic means that it is related to practical affairs: useful, but not necessarily true. From such a pragmatic perspective, social desirability in personality questionnaires is not necessarily problematic. Consider the following item from the International Personality Item Pool (Goldberg et al., 2006): "I complete my duties as soon as possible." Now imagine two job applicants: one of them "completely agrees" and the other one "completely disagrees" with the above item. Which one is better suited to the job? The first applicant is probably responding in a socially desirable manner, but at least knows what is likely to be desirable in the workplace, whereas the second applicant might not. A pragmatic test user can, thus, even benefit from a response style that makes good candidates look even better, especially given the empirical evidence that social desirability (as measured by social desirability scales) does not seem to compromise the predictive value of questionnaires (Ones, Viswesvaran, & Reiss, 1996).

A less common meaning of 'pragmatic' refers to pragmatics, a subfield of linguistics, studying the way language is used and understood. "Pragmatics is the science

E-mail: kenn.konstabel@ut.ee

of language seen in relation to its users" (Mey, 2001, p. 5). Questionnaire responses are examples of language use, so pragmatics might have something relevant to say. Or perhaps not—let us try to figure this out.

Evaluation as a Property of Words

A lexicographer might view evaluation as an attribute of words. From this angle, in addition to being symbols and referring to an object or a concept, words are like things that have attributes such as warmth (Asch, 1946), strength (Osgood, Suci & Tannenbaum, 1957), and why not also thickness, color, spin, and charm. Users of the lexicon can thus pick the word that has the attributes that most closely correspond to the meaning that they want to express. Thus, the larger one's vocabulary, the more concisely one can express one's ideas. For example, borrowing from Allport and Odbert's (1936) monograph of trait names, a politician might be quite truthfully described as pachydermatous or quixotic, but this does not mean that all politicians can be lined up on the dimensions of pachydermatousness and quixoticity. Common dimensions are a matter of empirical study and might be considerably more general (Caprara, Barbaranelli, & Zimbardo, 1997); specific words refer to common dimensions but have an additional, connotative nuance of meaning, which might be informative of the speaker's attitude towards the object.

Having two separate words for the same thing, as T. L. Kelley (1927) has noted, is "contaminating to clear thinking" (p. 64). Ironically, Kelley made the very same error that he was criticizing by calling it the 'jangle fallacy'—the name is just another name for synonymy (or another 'jangle', as Kelley might have said). Could he have said 'synonymy' instead of 'jangle fallacy'? Perhaps not that easily: 'jangle' comes off as something that one should avoid, especially when the purpose is clear thinking; 'synonymy', however, sounds like something rather pleasant. On the one hand, Kelley's main tenet was, of course, terminological clarity, which is of utmost importance. On the other hand, he was just one step away from raising the question of the evaluative and descriptive aspects of trait words: in the paragraph introducing the subsequently famous 'jangle fallacy,' he dismissed the difference between 'fearful' and 'cowardly,' as well as that between 'upright' and 'honorable,' as merely a creation of "literary ingenuity" (Kelley, 1927, p. 64). Nine years later, Allport and Odbert (1936, p. 32) disagreed:

> *Contrary to Professor Kelley's opinion, it is not literary ingenuity that is responsible for the distinctions here mentioned. Novelists take what they find in life. During the Great War it was by no means uncommon to find heroes who were fearful but not cowardly, and the record of the historical Puritan shows that often indeed he was upright but not honorable. Professor Kelley assumes that the criterion of a trait lies in its average occurrence in "the rank and file" rather than in the complex nature of single human beings.*

Thirty-odd years later, Peabody (1967) suggested that trait adjectives combine evaluative and descriptive aspects, and that, by using carefully orchestrated sets of adjectives, one can tell one from the other. For example, 'stingy' and 'thrifty' refer to the same general attitude towards money but are opposites on an evaluative dimension; the same can be said for 'extravagant' and 'generous.' This approach describes the average person in an average situation, but does not account for deviations from the average: individual differences in understanding words ('extravagant' may be evaluated as good by a few people who value extravaganza), situational and contextual differences ('hero' as used ironically as opposed to seriously; 'modest' as used to describe a work of art, or the author's evaluation of it), and, finally, historical changes. Stearns (1994) relates a story of a student who wrote in an examination paper that Columbus was heartily welcomed on his return to Spain, and, when asked, referred to a textbook stating that the explorer was given a "cool reception." Can we now say that the meaning of the word 'cool' has changed over time, and it is time to update the language used in the textbook? Not at all; this brings us to the question of homonymy, or, as professor Kelley would have said, the "jingle fallacy." Different meanings of 'cool' can peacefully coexist in the dictionary, and without knowing the context, there is no way to tell which one the speaker had in mind. The student had an imperfect grasp of the context, making it possible for her to interpret 'cool' as meaning 'fancy.' So, for a complete account of evaluation in questionnaire responses, we have to consider pragmatics.

Evaluation as an Aspect of Language Use

In his much cited 43rd passage of Philosophical Investigations, Wittgenstein (1953/2005) has written that the word's meaning can, in most usage cases, be clarified as its use in language. How much this aphorism clarifies things is another matter, but it is often used by those who think that meaning is related to how words are used (pragmatics) rather than fixed definitions (semantics). A vocabulary that contains only words that are free of evaluation and ambivalence, is, in principle, conceivable. But can one use that vocabulary without re-introducing both sins? An Irish scholar (Swift, 1726) described one such project: learned men in the Academy of Lagado decided to abolish all words and start using things instead. Even though the project's main rationale was public health[1], one might imagine that the professors could have also aimed for increased clarity of expression and decreased evaluativeness as side effects.

But could this actually work? Can the use of things instead of words, or, more generally, the use of only precisely defined symbols, get rid of ambiguity and evaluation? Imagine a member of the Great Academy of Lagado raising his hand with a big loaf of bread and pointing the other hand toward you. Does he want you to bring

[1] "For it is plain, that every word we speak is, in some degree, a diminution of our lunge by corrosion, and, consequently, contributes to the shortening of our lives" (Swift, 1726, part III, chapter V).

him bread or is he asking you to help him get rid of it? Is he offering the bread as a prize if you help him find the right road, or threatening to hit you with it if you do not stop chattering? If context does not make it clear, then we can say that his attempt at communicating was a failure.

Evaluation cannot be located at any specific level of language, but permeates all communication (Alba-Juez & Thompson, 2014), perhaps because we tend to communicate about things that matter to us, and, moreover, much of our early social life is devoted to learning how to evaluate things (Olsson, FeldmanHall, Haaker, & Hensler, 2018).

Interlude: On Academic and Festive Style

In a Festschrift, one is expected to do the scribing in a festive manner, or so I suppose. It is thus my hope that the gentle reader will generously pay only minimal attention to my occasional deviations from academic style, and my futile attempts at covering these up with citing the classics, such as Wittgenstein (1953) or Swift (1726). And conversely, I hope that full attention will be paid to the *raison d'être* of this chapter, namely, wishing Professor Allik a very happy birthday!

An Atomistic Model of Communication

In the real world, personal descriptions are often complex, and sometimes even attempt to be exhaustive. For example, Georges Frédéric Parrot (1819, p. 6), in the introduction to his *Entretiens sur la Physique*, describes all the main characters of the book, and among others, a certain Monsieur de R, as a:

> *… man of letters, loveable by his extreme bonhomie and a touch of enthusiasm that he spreads on everything that he says, a friend of the table and of women, easy by temperament, serious by principles, passionate to know the daily chronicle, an author of a profound opus, hardy and eloquent in finances.*

The author is considerably less verbose about himself, stating that he "does not merit the honour of being named" (Parrot, 1819, p. 7). Both descriptions are socially desirable: the first one gives the impression of a neutral catalogue of attributes, but, in fact, is mostly positive; the traits that could be perceived as negative are, in the fashion of true bonhomie, presented from an amiable angle ("a *friend* of the table and of women"; "*passionate to know* the daily chronicle"). One could say that the first description exemplifies the value of *benevolence,* whereas the second, seemingly less benevolent, reflects the values of *modesty* and *humility* (cf. Schwartz, 1992).

In this essay, I shall confine myself to less complex descriptions, often as short as a single word, or a single number indicating the degree of agreement with a short sentence. The complexities of the communicative situation can thus be reduced to a few elementary components, such as those described by Jakobson (1960): the *addresser* sends a *message* to the *addressee,* the message uses a *code,* and presumes a *context* (in Jakobson's interpretation, everything that is outside of the text, i.e., including the external reality), and the *contact.* As proposed by Jakobson (1960), and later systematized by Klinkenberg (1996), each of these components corresponds to a function of language: for example, *context* is coupled with the *referential* function, the capability of words to refer to something outside of the utterance. Can social desirability be located within a single component of the communicative situation, and, thus, to a single function of language? Let us consider these issues one by one.

Social Desirability and the Referential Function of Language

In the above cited paper, Jakobson (1960) seems to view the referential function as the most obvious one, paying it the least attention. But unpacking the "context," as the term is used by Jakobson, reveals a hidden complexity: it includes the immediate surroundings of the communicative situation and previous turns in conversation, as well facts about similar conversations and the state of art in the world.

Consider a person completing a personality questionnaire. The questionnaire's author hopes that respondents are describing themselves as they really are, regardless of evaluation. But wait a second! A peaceful and benevolent citizen is socially more desirable ("better") than an axe murderer, so how can one expect their respective descriptions to be free of evaluation? In personality psychology, evaluation is usually discussed as something to be gotten rid of. Willem Hofstee (2003), however, is one of the few to see things differently: most people actually do have socially desirable qualities; if one quantifies the overall amount of "desirability" of respondents as traits or items weighed by their respective social desirability values, there are some individual differences but, overall, most people fall on the "positive" side, with only a few receiving a "negative" or socially "undesirable" score. It is important to note that the question of "being" socially desirable is orthogonal to "reporting" oneself as desirable; the two may in some cases be even contradictory—for instance, by 8 years of age, most children become aware of the detrimental effect that lack of modesty may have on social evaluation (Banerjee, 2000).

Beauvois and Dubois (2009) go even further, asserting that traits are inherently evaluative, and rather than being descriptions of the average behavior of a person, perceived traits inform us about how one can behave towards a person. For example, an aggressive person may be someone who "yells at others," but he is also (and perhaps in a more useful sense) "someone you avoid provoking" (Mignon & Mollaret, 2002). Beauvois and Dubois (2009) describe the main evaluative dimensions in personal perception as social desirability and social utility, which are similar to the egoistic and

moralistic biases described by Paulhus and John (1998). The approaches of Paulhus and Beauvois differ, however, in their emphasis: the first would view evaluation as a bias that corrupts the truth in personality judgments; the second contends that personality judgments are inherently evaluative and contain 'truth' only to the extent that they help with the main purpose of reaching a useful evaluation. For the present purpose, it suffices to say that both positions make sense, but neither seems to tell the whole story. One practical implication is that the idea of writing evaluatively neutral items (Bäckström & Björklund, 2014) can succeed only to the extent that evaluation can be treated as bias and the content to be asked about is inherently neutral. In the modal case, one can make an item more or less neutral, but not completely so: Bäckström and Björklund's (2014) example of "I am exacting in my work" being "neutralised" as "Continue working with a task so that every small detail is right" does not *just* remove bias, it also changes the content.

In addition to describing one's current behavior or current self, respondents can also describe what they believe they would do in a given situation in the future. For example, job applicants may, regardless of their ordinary levels of conscientiousness and agreeableness, believe that they would work conscientiously in every task and be agreeable with every colleague in their new job. Although the questionnaire asks about their usual behavior, they may partly describe what they imagine (or 'plan,' to add another 'jangle') to be true in the future. This need not be a lie, let alone a deliberate one: respondents may perceive their future behavior as workers as highly *relevant* (cf. Sperber & Wilson, 1995) to the questions at hand—after all, testing is carried out to evaluate their suitability for the job, not, for example, to get to know about their ideas about modern art or social policy (unless these are directly related to the job). That is, rather than answering the literal question, respondents may take the asker's purpose into account and respond with information they perceive as being expected from them in their new job. This is what we routinely do in everyday life: for instance, one would ordinarily not interpret the utterances such as "Can you pass me salt?" or "Have you seen my car key?" as literal questions. Finally, the respondent's attempt to provide information that is relevant for the purpose of testing (for example, describing one's "work self" rather than "home self" when applying for a job) may be helpful for the test user: for instance, Lievens and colleagues (2008) have shown that contextualized questionnaires predicted work performance slightly better than overall personality questionnaires.

The Emotive Function: Condition of the Sender

Klinkenberg (1996) defined this function as the extent to which a message reveals the condition of its sender; consequently, he has argued that it should be called "expressive" rather than "emotive."

The influence of temporary moods on responses to personality items is an obvious example of the emotive (or expressive) function playing a role in socially desirable

responding. Consider a person who, uncharacteristically, has recently made a bad decision that she regrets, and is now given the Balanced Inventory of Desirable Responding (Paulhus, 1984) to fill in. She responds with a "2" (on a 7-point scale) to item no. 5 ("I always know why I like things") because she is not sure why she made the aforementioned decision. For item no. 11 ("I never regret my decisions"), she clicks "1", expressing the lowest degree of agreement, and for item no. 15 ("I am a completely rational person"), she regrets not having an even more strongly disagreeing response option. In sum, temporary moods can make people feel more or less optimistic about themselves, and this has an obvious impact on the social desirability of their responses.

Another possibility is reviewed by Leary (2007): having a certain belief about oneself may be a motivation in its own right. It is nice to believe that I have many wonderful qualities, and, at the same time, am immune to all the sins one can imagine. Interestingly, people are partly aware of their self-serving biases: they can tell, with some accuracy, on which traits they have probably rated themselves more positively (or negatively) than their friends' ratings of them (Bollich, Rogers, & Vazire, 2015).

The Conative Function: Making the Addressee Do or Believe Something

Orders and prohibiting signs are prototypical examples of the conative function, as put forward by Jakobsen (1960), but Klinkenberg (1996) also mentions more subtle forms, such as advertising and persuasion. A prototypic example from questionnaire-related behavior is deliberate impression management—describing oneself in a more favorable manner, despite knowing that this is not the whole truth, and, occasionally, not true at all. Paulhus (1984) has distinguished impression management from self-deceptive enhancement; from a present perspective, true self-deception would be primarily classified under the expressive function (see above). However, the difference is not always that clear, and it should not be taken for granted that, for instance, an impression management (IM) scale measures deliberate impression management. As an example, one of the items in the IM subscale of the BIDR (Paulhus, 1984) is "I never swear;" one can imagine a heavy swearer trying to cover up his true self in a questionnaire and make the test interpreter believe that he never swears, but equally likely is someone who thinks this: "Everyone has uttered a swearword at least once in their life, so that's not what they're after. They must have meant to ask whether I am heavy swearer, and I am not, so I suppose I can tick 'agree' or 'strongly agree', depending on how heavy is heavy." There is nothing lamentable about this interpretation – we do it all the time in everyday conversations. Someone might, for instance, say, "Why are you nagging at me all the time?," and by "all the time" the person would usually not mean "with no interruption in nagging lasting more than 200 milliseconds," but rather something like "at least once a day" or simply "too much." It is thus no surprise that IM scales have not been found to be of great use in detecting deception in personality questionnaires (e.g., Ones et al., 1996). There are indications, however, that high IM may be a sign of other-oriented self-control, and not deceiving others (Uziel, 2013).

The Phatic Function: Keeping the Contact

The phatic function (Jakobson, 1960) is focused on contact: establishing and keeping contact and making sure one still has it. The prototypic examples are greetings, talk about the weather or climate, and other "conversations of no or of little importance," as Klinkenberg (1996) has concisely put it. Instead of weather, one could use any topic of general interest, as in the following dialogue:

A: The vast majority of men are truthful and dependable.
B: Exactly my idea; trust in others is what makes people happy.
A: Happiness is one of the primary goals of life.
B: I can't agree more.
A: No one cares much about what happens to you.
B: That's a sad fact of life.

The first two lines attributed to person A are borrowed from an acquiescence questionnaire by Couch and Keniston (1960); the third one is from Edwards' Social Desirability Scale (Paulhus, 1991). Agreeing with an item that most people agree with may be good for keeping conversation going, but it is not, *sensu strictu*, informative. One cannot but notice that many of the most "agreed-upon" items are highly evaluative; for example, A's opening line is similar in content to the items of the NEO PI-R (Costa & McCrae, 1992) Trust subscale of Agreeableness (A1), which is positive (although not extreme) on social desirability, as shown by Konstabel, Aavik, and Allik (2006).

In some cases, thus, acquiescence and social desirability may partly overlap: the respondent may agree with an item because this is an easy answer to give, and the item is evaluatively positive. Because evaluation is usually quickly perceived, this may extend to rejecting negative items as well; thus, one cannot expect balanced keying—the standard strategy for combatting acquiescence—to work with the "phatic" type of "yea/nay-saying."

Peabody (1966) has described another type of acquiescence that may seem contradictory to social desirability: agreeing strongly and indiscriminately (or even, as Peabody put it, "simple-mindedly") with items such as "Obedience and respect for authority are the most important virtues children should learn." This idea is generally valued in some societies and generally denounced in others, but even in the latter, it may be viewed as desirable in certain subgroups.

The Poetic Function: Message for its Own Sake

The poetic function focuses on the message for its own sake, which often means the form of the message: how something is expressed, which words from among countless possibilities are chosen, and how they are assembled. Apart from poetry, this function is characteristic of slogans, proverbs, and ritualistic language (Klinkenberg, 1996).

In a personality questionnaire, one would ordinarily want to avoid the poetic function, so as to not catch the respondent's eye too much; if it does, it is often

perceived as awkward. One of my favorite examples is an item from the Parenting Sense of Competence Scale by Johnston and Mash (1989): "I go to bed the same way I wake up in the morning, feeling I have not accomplished a whole lot." There are numerous ways to ask the same question without mentioning bed. One could, moreover, spend a day writing parodic paraphrases of this item but the conclusion one would reach by bedtime would still be the same: The Hidden Brain Damage scale by Wegner and colleagues (1979) should be an introductory reading for everyone who understands irony and still wants to construct a psychometric scale.

One could say that the ideal personality questionnaire item is as "un-poetic" as possible: attention to the form of the message is justified as long as it can make the message less ambiguous, easier to understand, and more "natural-sounding" to the respondent. Occasional lapses into poetry are usually a sign that something is out of place: for instance, if the original item is "I have never felt joy about other people's failures" and one encounters the word "rejoice" in a back-translation[2]. These are the precious moments that earnest item writers would better keep for their own private entertainment: using common, everyday language in items makes it less likely that respondents will be distracted by irrelevant poetry.

The Metalingual Function: What Do You Mean by "Metalingual"?

According to Jakobson (1960), metalingual function means communicating about the code (language), as in explaining the meaning of a word, or asking for such clarifications. This function is evident in instructions to the respondent, which often state that there are no "good" or "bad" answers, that is, any answers are evaluatively neutral. Apart from such instructions, one would ordinarily avoid lengthy definitions in a questionnaire, and one would try to phrase items so as to avoid the necessity of any metalingual questions.

Autocommunicative Function

Metaphorically, Jakobson's (1960) model resembles the telegraph: there is a sender who uses a code, and a channel to deliver a message to the addressee. In human beings, matters are further complicated by autocommunication, which is, at least potentially, an aspect of every communicative act (Lotman, 1990). This may be relevant in responding to a personality questionnaire: a respondent may take an item as a reminder of something that he might want to do or pay attention to in the future. The respondent may read an item such as "I don't gossip about other people's business," and this may remind him, for instance, about the last time he did gossip, and the embarrassing consequences that followed, and may lead to a renewed decision to reduce his gossiping behavior to a reasonable minimum. The response is thus moved

[2] The back-translated item was "I have never rejoiced other people's failures," whereas the original translation was considerably less poetic ("Ma ei ole kunagi teiste ebaõnnestumise üle rõõmu tundnud").

in a socially desirable direction, but is not necessarily misleading, when it comes to the prediction of behavior.

Conclusions

Social desirability is not just about the meaning of words or sentences; it is about the whole communicative situation of personality description. Using Jakobson's (1960) model of communication as a starting point, various aspects of socially desirable responding were reviewed. The desirable response may reflect a desirable quality of the person, or the respondent's understanding that desirable qualities are relevant to the purpose of communication; it may reflect the respondent's motivated beliefs about the target, or the respondent's goal to make the test user believe something socially desirable about the target. A socially desirable response may just be an answer that is easy to give, informing us mostly about the fact that the respondent has processed the question and produced a normative answer. Finally, it can serve autocommunicative purposes, reminding respondents of a socially desirable belief or action that they would like to have or perform. These communicative effects can all result in social desirability, that is, respondents agreeing with socially desirable items and disagreeing with socially undesirable items.

References

Alba-Juez, L., & Thompson, G. (2014). The many faces and phases of evaluation. In G. Thompson & L. Alba-Juez (Eds.) *Evaluation in context* (pp. 3–23). Amsterdam: John Benjamins.

Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs, 47*, i–171.

Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal Psychology, 41,* 258–290.

Bäckström, M., & Björklund, F. (2014). Social desirability in personality inventories: The nature of the evaluative factor. *Journal of Individual Differences, 35*, 144–157.

Banerjee, R. (2000). The development of an understanding of modesty. *British Journal of Developmental Psychology 18,* 499–517.

Beauvois, J.-L., & Dubois, N. (2009). Lay psychology and the social value of persons. *Social and Personality Psychology Compass, 3/6,* 1082–1095.

Bollich, K. L., Rogers, K. H., & Vazire. S. (2015). Knowing more than we can tell: People are aware of their biased self-perceptions. *Personality and Social Psychology Bulletin, 41,* 918–929.

Caprara, G. V., Barbaranelli, C., & Zimbardo, P. G. (1997). Politicians' uniquely simple personalities. *Nature 385,* 493.

Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.

Couch, A., & Keniston, K. (1960). Yeasayers and naysayers: Agreeing response set as a personality variable. *The Journal of Abnormal and Social Psychology, 60,* 151–174.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. C. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality, 40,* 84–96.

Hofstee, W. K. B. (2003). Structures of personality traits. In T. Millon & M. J. Lerner (Eds.), *Handbook of psychology: Personality and social psychology, Vol. 5* (pp. 231–254). Hoboken, NJ: John Wiley & Sons Inc.

Jakobson, R. (1960). Closing statement: Linguistics and poetics. In T. A. Sebeok (Ed.), *Style in language* (pp. 350–377). Cambridge, MA: The MIT Press.

Johnston, C., & Mash, E. J. (1989). A measure of parenting satisfaction and efficacy. *Journal of Clinical Child Psychology, 18,* 167–175.

Kelley, T. L. (1927). *Interpretation of educational measurements.* Yonkers-on-Hudson, NY: World Book Company.

Klinkenberg, J.-M. (1996). *Précis de sémiotique générale.* Bruxelles: De Boeck.

Konstabel, K., Aavik, T., & Allik, J. (2006). Social desirability and consensual validity of personality traits. *European Journal of Personality, 20,* 549–566.

Leary, M. (2007). Motivational and emotional aspects of the self. *Annual Review of Psychology, 58,* 317–344.

Lievens, F., De Corte, W., & Schollaert, E. (2008). A closer look at the frame-of-reference effect in personality scale scores and validity. *Journal of Applied Psychology, 93,* 268–279.

Lotman, Y. M. (1990). *Universe of the mind: A semiotic theory of culture.* (A. Shukman, Trans.) London: I. B. Tauris.

Mey, J. L. (2001). *Pragmatics: An introduction* (2nd ed.). Oxford, UK: Blackwell.

Mignon, A., & Mollaret, P. (2002). Applying the affordance conception of traits: a person perception study. *Personality and Social Psychology Bulletin, 28,* 1327–1334.

Olsson, A., FeldmanHall, O., Haaker, J., & Hensler, T. (2018). Social regulation of survival circuits through learning. *Current Opinion in Behavioral Sciences, 24,* 161–167.

Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). The role of social desirability in personality testing for personnel selection. *Journal of Applied Psychology 81,* 660–679.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning.* Urbana, IL: University of Illinois Press.

Parrot, G. F. (1819). *Entretiens sur la Physique.* [*Conversations on Physics*]. Tome Premier (avec 2 planches). Dorpat: J. C. Schünmann, Imprimeur de l'Université impériale de Dorpat.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46,* 598–609.

Paulhus, D. L., & John, O. P. (1998). Egoistic and moralistic biases in self-perception: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality, 66*, 1025–1060.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of social psychological attitudes, Vol. 1. Measures of personality and social psychological attitudes* (pp. 17–59). San Diego, CA, US: Academic Press.

Peabody, D. (1966). Authoritarianism scales and response bias. *Psychological Bulletin, 65,* 11–23.

Peabody, D. (1967). Trait inferences: evaluative and descriptive aspects. *Journal of Personality and Social Psychology, 7,* 1–18.

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. *Advances in Experimental Social Psychology, 25*, 1–65.
Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition,* (2nd ed.). Oxford: Blackwell.
Stearns, P. (1994). *The American cool.* New York: NYU Press.
Swift, J. (1726). *Travels into several remote nations of the world. In four parts. By Lemuel Gulliver, first a surgeon, and then a captain of several ships.* London, UK: Benjamin Motte.
Uziel, L. (2013). Impression management ("lie") scales are associated with interpersonally oriented self-control, not other-deception. *Journal of Personality, 82,* 200–212.
Wegner, D. M., Vallacher, R. R., & Gilbert, C. (1979). The hidden brain damage scale. *American Psychologist, 33,* 192.
Wittgenstein, L. (2005). *Filosoofilised uurimused [Philosophische Untersuchungen].* (A. Luure, trans.) Tartu, Estonia: Ilmamaa. (Original work published in 1953.)

PART III: EMPIRICAL CONTRIBUTIONS

8 OUTLIERS IN PERCEPTUAL AVERAGING: OVER- OR UNDERWEIGHT?

AIRE RAIDVEE

Our visual system can form instant summary representations of object ensembles. Whether this process is global or involves sub-sampling (or differential weighting of elements) is not clear. Recent evidence suggests that items with values further from the ensemble mean or the decision boundary (i.e., outliers) are ignored or down-weighted when computing the average. The present study tests whether outlier down-weighting is a universal property of perceptual averaging or whether it depends on observers' prior knowledge of the location of the decision boundary. In the present study, observers were asked to compare the average tilt of eight Gabor patches to the tilt of a single reference. The tilt of the reference was either kept constant across all trials, or randomized. The results show that all evidence was weighted equally in both conditions. Therefore, it seems that outlier down-weighting is neither universal nor invariably caused by prior knowledge of the stimulus space.

With deepest gratitude to my beloved mentor, Jüri Allik, whose wisdom, generosity, and great sense of humor I have had the fortune to enjoy for, by now, half of my life.

Introduction

The concept of perceptual averaging is appealing for describing the mechanism with which the visual system handles its capacity limits. Indeed, our perception is able to instantly form summary representations of large ensembles of various types of object dimensions, ranging from lengths of lines (Miller & Sheldon, 1969) and sizes of circles (Allik, Toom, Raidvee, Averin, & Kreegipuu, 2013; Chong & Treisman, 2003) to the lifelikeness of object groups (Leib, Kosovicheva, & Whitney, 2015). This idea of perceptual averaging implies that all or most of the to-be-averaged elements are actually processed. Evidence on whether this is the case is still mixed. On the one hand, some data suggests that the visual system automatically prefers summary representations of ensembles over those of detailed individual objects, and even lacks conscious access to the latter (Allik, Toom, Raidvee, Averin, & Kreegipuu, 2014; Ariely, 2001; Chong & Treisman, 2003, 2005). Moreover, perceptual averaging has claimed to not only be automatic, but compulsory, involving all available items (Morgan, Hole, &

I would like to thank all participants for their patience and time and Daryl Fougnie and Weiji Ma for the inspiring discussions on outlier treatment in perceptual averaging.

E-mail: aire.raidvee@ut.ee

Glennerster, 1990; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001), irrespective of attentional deployment (Alvarez & Oliva, 2008). On the other hand, it has been shown that, rather than positing a global (and unlimited) process in judging ensemble statistics, simpler mechanisms of focused (limited) attention strategies could explain the data. Myczek and Simons (2008) successfully simulated empirically expected accuracy levels on the basis of the assumption that only a few random items were included in the judgment on the global mean. Zero perceptual noise was assumed in their simulations, which is why the resulting efficiency estimate of only a few items likely falls short of reality (Im & Halberda, 2013). The trade-off relationship that exists between efficiency (i.e., sampling rate reflecting how many available items are effectively processed) and perceptual noise is exactly what makes it a challenge to determine whether, and to what degree, ensemble processing is, in fact, global. This means that an agent that accounts for all elements but with a high degree of internal noise may yield estimates of a similar accuracy to one that noiselessly accounts for only a few items. Thus, the resulting estimates of their respective processing efficiencies (based on accuracies) would be similar, despite the different processing strategies.

In other words, telling the effects of inattention and perceptual noise apart has been a challenge. Here, the study of outlier treatment in ensemble processing has proven useful. Since outliers have a pronounced effect on the estimate of the average, it is possible to theoretically judge the degree to which they contribute to the summary statistic. The most extreme case of differential weighting would be subsampling, meaning that some items' weight would be zero (i.e., they would be ignored).

The particular weight of each item has been found to be related to its attentional saliency (de Fockert & Marchant, 2008; Mareschal, Morgan, & Solomon, 2010). Yeshurun and Carrasco (1998) showed that attentional saliency increases the spatial resolution of the internal representation of the attended-to-location. It has been, thus, speculated that items may be weighted by their perceived reliability or precision (Alvarez, 2011), but the question of whether, how, and to what degree differential weighting is actually applied remains open.

Some recent evidence suggests that item weights depend on their location in the stimulus space—observers tend to ignore extreme data, i.e., they down-weight outliers similarly to statisticians (de Gardelle & Summerfield, 2011). One study claimed that this outlier down-weighting improves decision accuracy in the face of neural noise (Li, Herce Castañón, Solomon, Vandormael, & Summerfield, 2017).

The current experiments were directly inspired by a study by de Gardelle and Summerfield (2011), which used a task where participants were requested to discriminate the average color (ranging from red to blue), or average shape (ranging from a square to a circle) of eight elements. The comparison was always made relative to the mean location of the stimulus space (which was kept constant throughout the experiment). Given the constancy of the stimulus space boundaries as well as the reference value, it seems plausible that participants realized that "zooming in"

around the critical area, closer to the reference value, offered bigger gains (note that they received feedback after each trial). Therefore, it seemed worthwhile to test the treatment of feature values further removed from the reference in a situation where a) the stimulus space is not bounded (which is conveniently the case with circularly dimensioned features such as angles) and b) the reference value is dynamic, rather than stationary, across trials.

To conclude, the goal of the present study is to test the generalizability of outlier down-weighting (or, so-called robust averaging) to situations where participants cannot build prior knowledge on the bounds of the feature space nor the location of the reference within that feature space. The second, and more important, aim is to address some of the research questions that the study entailed, such as, the degree to which outlier weighting is related to attentional saliency, and whether weighting patterns are optimal, especially in case of a limited capacity, noisy integrator.

Methods

Participants

The experiments were carried out in July 2016 under the approval of the Ethics Review Board of University of Tartu. All participants gave written informed consent and received a small amount of financial compensation for their time. All participants (recent graduates of a high school in Tartu) but one (the author of this study) were naïve to the purposes of the experiments.

Five adults (two females; four aged 18–19 years and one 40 years) participated in Experiment 1, and six adults (one female; all aged 18–19 years) participated in Experiment 2 (four of them had also taken part in Experiment 1). One participant in Experiment 1 and one in Experiment 2 were excluded from the analyses because their responses were at chance level.

Apparatus

The stimuli were generated on a standard LCD monitor (frame rate 60Hz) with the help of Matlab and the Psychophysics Toolbox (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) and CircStat Toolbox (Berens, 2009).

Task and Procedure

Participants were asked to judge the global average orientation of the presented set of eight Gabor patches relative to the angle of a single test patch. All stimuli were presented at a viewing distance of 60 cm, without a headrest, in a darkened room, on a rectangular gray background of luminance 57 cd/m^2, subtending 34.4 × 19.9 degrees of visual angle.

Each trial started with a fixation cross presented for 800 ms. The test display contained a cloud of eight Gabor patches (with a phase of 0°, 2.3 cycles per degree; the value of the spatial constant of the Gaussian hull function was 11, with a Michelson contrast of 75%—luminance ranging from 20 to 139 cd/m^2, 1 degree of visual angle per grating). The Gabor patches were positioned randomly in the central annulus of the screen (diameter 10.6 degrees of visual angle) in a way that prohibited the centers of any two Gabor patches from being closer than 1.2 degrees of visual angle. The Gabor patches were presented for 200 ms.

The orientation values of the Gabor patches were drawn from a von Mises distribution, with a standard deviation of 7°, 14°, or 21°. The generating value of the distribution mean was set apart from the reference angle by -10°, -5°, 5°, or 10° degrees (randomized separately for each trial). The only detail that differed between Experiments 1 and 2 was the direction of the reference angle—in Experiment 1 the angle was constant across trials within a single participant (randomized across participants), whereas in Experiment 2 the angle was always randomized, on a trial-by-trial basis, from a uniform distribution across the entire circle.

Following the test display, a response screen with a central Gabor patch tilted at the reference angle was presented. Participants were requested to indicate with a mouse click whether they judged the average angle of the previous test display to be either clockwise or counter-clockwise relative to the reference angle. The response phase of each trial was self-paced and feedback on response accuracy was provided. A schematic view of the task sequence of each trial is shown in Figure 8.1.

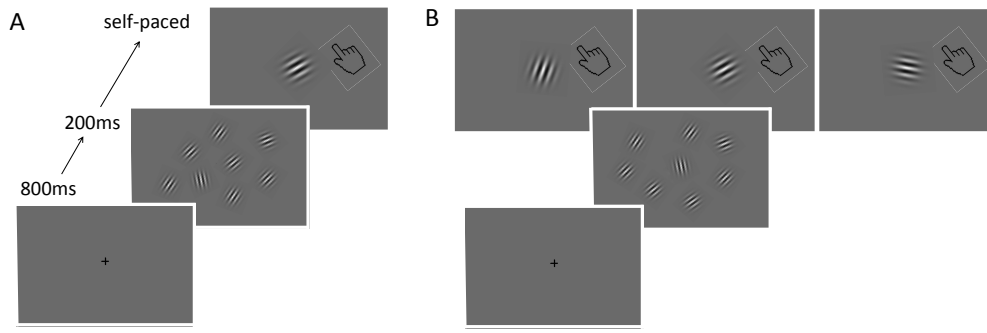


Figure 8.1 Trial displays. The stimulus set followed by the response screen in Experiment 1, with a constant reference angle (panel A), and a response screen with a randomized reference angle in Experiment 2 (panel B).

One experimental session started with 30 training trials, which were followed by 720 randomly ordered test trials divided into 10 blocks (participants could take breaks between the blocks). Participants completed two experimental sessions (one per each of the two experiments) over two consecutive days.

Statistical Analyses

For each participant and each trial, the circular distance between the circular average angle θ of the eight Gabor patches and the reference angle β was computed, and the accuracy of the response was determined. The angular orientations of the Gabor patches were divided into eight bins and the binary accuracy of responses was regressed on the relative weights of each bin of the stimulus space (using probit regression). The binnings were carried out in two ways. Firstly, the angles were ranked within each trial, such that the first bin contained angles that were the most counter-clockwise within a particular trial, and the last (eighth) bin contained the most clockwise angles (i.e., the ranks were not locked with respect to the reference value). This analysis rationale followed that which was applied by de Gardelle and Summerfield (2011). Secondly, the angles were divided into eight bins based on their location with respect to the reference value (from -45° to 45° from the reference, with cut points between bins at -45°, -31°, -18°, -6°, 0°, 6°, 18°, 31°, and 45°). In both types of analysis, only trials where all test angle deviations from the reference were within 45° were included (thus, of all trials, 85.6% were included for the random reference, and 84.8% for the constant reference, conditions).

Analyses were carried out in Matlab and *R* (R Core Team, 2018).

Results

The two experiments were designed to test whether outlier down-weighting is a general property of perceptual averaging. Based on previous research (de Gardelle & Summerfield, 2011; Li et al., 2017), it was expected that the results from Experiment 1 (with the reference angle fixed across the trials) would reflect robust averaging (outlier down-weighting). It was reasoned that robust averaging would be observed because people were able to learn, over the course of trials that the location of the reference angle (i.e., the decision boundary) as well as the bounds of the stimulus distribution were stable, and could, therefore, attribute greater significance to the area of the stimulus space closer to the reference. In Experiment 2, the reference angle was randomized (from a uniform distribution over the circle) on trial-by-trial basis in order to make it impossible to build any knowledge on the location of the reference or on the bounds of the stimulus space. It was thus expected that people would apply equal weights across the entire stimulus space (which would be the optimal strategy).

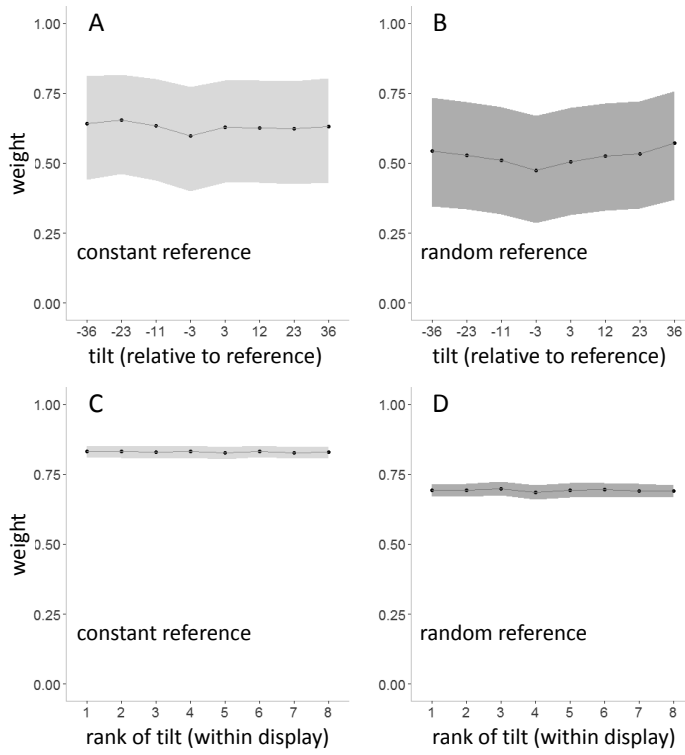


Figure 8.2 Parameter estimates from probit regressions from Experiment 1, with a fixed reference (panels A and C), and Experiment 2, with a reference randomized across trials (panels B and D). Parameter estimates reflect the weights (on a probability scale) of Gabor patches (belonging to the designated categories) in predicting the correct choice. In the upper row of panels (A and B), Gabor patches were binned based on their relative distance from the reference. The labels on the *x*-axis reflect the mean angle of Gabor patches in the respective bin in degrees. In the lower row of panels (C and D), Gabor patches were binned according to their rank within a single trial display (irrespective of their distance from the reference). The labels on the *x*-axis reflect the rank of the tilt of a Gabor patch within a single display from the most counter-clockwise to the most clockwise. Shaded areas represent the standard errors of the mean.

The weights attributed to different areas of the stimulus space are reflected in Figure 8.2. The location of the reference angle is always taken to be at 0°, thus, on all panels of Figure 8.2, the outlying stimulus values are depicted by the left- and right-most bins. Parameter estimates on the *y*-axis reflect the weights (on a probability scale) of the Gabor patches (belonging to the designated bins) in predicting the correct choice. The higher the weight assigned to a certain bin, the greater the importance of the Gabor patches belonging to that bin in predicting an accurate response. Magnitudes of the weight estimates are proportional to the overall rates of correct responses, which are 70.5% and 62.9% in the constant and random reference conditions, respectively (with 95% binomial confidence intervals of [69.0%, 71.9%] and [61.3%, 64.5%], respectively).

Contrary to expectations, the results from Experiment 1 are not reflective of robust averaging (Figure 8.2; panels A and C). In fact, all evidence was weighted equally in both Experiments (in line with the expectation for Experiment 2). In Experiment 2, there is a mild trend towards outlier upweighting in case of random reference values. This trend is detected by an analysis based on the binned locations relative to the reference (see Figure 8.2 panel B), but not by analyses based on ranks of tilt within a particular display.

Discussion

Most of the (few) studies conducted to date on outlier weighting in visual perception have found evidence of outlier discounting. For example, in his (probably the first modern) paper on summary statistics in perception, Spencer (1961) found that, in judging the height of bars, the outlying value was always down-weighted. Another finding from low-level vision was that items with feature values further away from the center of the parameter space tend to get down-weighted (de Gardelle & Summerfield, 2011). Also, faces with emotional expression different from the rest were ignored in judging the summary representation of the set (Haberman & Whitney, 2010). A recent study argues that, in the face of neural noise arising during decision-making, outlier down-weighting is protective, as it improves decision accuracy compared to when all items are weighted equally (Li et al., 2017).

Unlike in visual perception, many examples of outlier overweighting are known from the cognitive domain, such as the availability heuristic by Tversky and Kahneman (1973), where people often attribute great importance to unusual events. It has also been found that, in mean estimates based on numerical symbols (presented one-by-one), people tend to over-weight more extreme numerical values (Spencer, 1961; Spitzer, Waschke, & Summerfield, 2017).

Contrary to previous findings from the perceptual domain by de Gardelle and Summerfield (2011) and Li et al. (2017), the results of the present study point towards non-robust averaging in orientation. On the assumption that the current result is not a spurious finding nor a technical artefact, it does follow the results of a study by Raidvee and Fougnie (2017), which reliably showed outlier upweighting in a similar display with slightly different parameters of the von Mises distribution—participants' reports on the display average systematically deviated in the direction of the outlier. It was shown that up-weighting was related to the outlier's attentional saliency. Several other studies on vision have demonstrated that the informational weight of display elements is related to their noticeability (Albrecht & Scholl, 2010; de Fockert & Marchant, 2008; Mareschal et al., 2010). Given that the angular distributions of the Gabor patches were very similar in the current study and that of Li et al. (2017)[1], with the primary

[1] The similarities between the experimental setups and angular distributions of the stimuli used in the present study to the ones used in Li et al. (2017) were fortunate but coincidental, as the current data were already collected when Li et al. (2017) was published.

difference being in the spatial positioning of the stimuli (randomly in an annulus vs. circularly around the central fixation point), it is conceivable that differences between the present and previous results may be related to (in)attention. Perhaps in a circularly positioned stimulus array, items with extreme feature values go unnoticed more often than when the stimuli are arranged in a cloud (as the latter may facilitate outlier pop-out more than the former). Even if a flat rate of inattention is assumed across the entire feature space, an item with a feature value further from the mean going unnoticed will have a larger effect on estimates, also because more extreme values are less frequent in the (circular) normal distributions used in the present study and in the study of Li et al. (2017).

It seems that the answer to the initial driving question for this study—does outlier down-weighting occur mainly because people learn the bounds of the feature space relative to the reference—is "no". Outlier treatment across the fixed and variable reference conditions was similar in both this and the study of Li et al. (2017). The new question is why outlier treatment differed across these two studies—outlier down-weighting in the latter vs. equal weighting in the current study.

Li and colleagues (2017) argue that, in a noisy integrator (such as human visual perception), robust averaging (outlier down-weighting) is protective against late noise. In any case, this does not seem to be universally so, especially given the relative similarity between the mathematical properties of the stimulus in the study of Li et al. (2017) and the ones used in the present study. While it is beyond doubt that, in a noiseless agent, the most optimal strategy would be to weight all items equally, it was shown by van den Berg and Ma (2012) that, even in the presence of noise, in the same task used in de Gardelle and Summerfield (2011), the optimal strategy was to weight all observations equally.

Limitations and Conclusions

Probably the main limitation of this study is related to its technical aspects. For example, it is not clear whether a generalized linear modeling approach is best suited for the analysis of circular dimensions. While it is true that linear mapping of angular data does create distortions, it is unlikely, as Li et al. (2017) also point out, that these distortions occur within a narrow sector of [-45°, 45°] in which angular data can be roughly linearly approximated. Nevertheless, other issues like the fact that feature values across different bins are correlated, remain. Based on preliminary simulations, it also seems that the probit regression is more powerful in detecting down-weighting than upweighting, but, presently, this is highly speculative.

In conclusion, it remains an interesting question which factors are facilitative of differential outlier weighting, and to what degree different weighting schemes are optimal in a noisy perceiver. Better understanding of outlier treatment would shed light on whether perceptual averaging is a global process or involves sub-sampling—a question relevant not only from the viewpoint of quantifying attentional scope but for the more general understanding of our mental architecture.

References

Albrecht, A. R., & Scholl, B. J. (2010). Perceptually averaging in a continuous visual world: Extracting statistical summary representations over time. *Psychological Science, 21*, 560–567.

Allik, J., Toom, M., Raidvee, A., Averin, K., & Kreegipuu, K. (2013). An almost general theory of mean size perception. *Vision Research, 83*, 25–39.

Allik, J., Toom, M., Raidvee, A., Averin, K., & Kreegipuu, K. (2014). Obligatory averaging in mean size perception. *Vision Research, 101*, 34–40.

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences, 15*, 122–131.

Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science, 19*, 392–398.

Ariely, D. (2001). Seeing sets: representation by statistical properties. *Psychological Science, 12*, 157–162.

Berens, P. (2009). CircStat: A MATLAB Toolbox for Circular Statistics. *Journal of Statistical Software, 31*, 1–21.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 433–436.

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research, 43*, 393–404.

Chong, S. C., & Treisman, A. (2005). Statistical processing: computing the average size in perceptual groups. *Vision Research, 45*, 891–900.

de Fockert, J. W., & Marchant, A. P. (2008). Attention modulates set representation by statistical properties. *Perception & Psychophysics, 70*, 789–794.

de Gardelle, V., & Summerfield, C. (2011). Robust averaging during perceptual judgment. *Proceedings of the National Academy of Sciences* of the *United States* of America*, 108(32)*, 13341–13346.

Haberman, J., & Whitney, D. (2010). The visual system discounts emotional deviants when extracting average expression. *Attention, Perception, & Psychophysics, 72*, 1825–1838.

Im, H. Y., & Halberda, J. (2013). The effects of sampling and internal noise on the representation of ensemble average size. *Attention, Perception, & Psychophysics, 75*, 278–286.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3. *Perception*, *36*, 1–16.

Leib, A. Y., Kosovicheva, A., & Whitney, D. (2015). Fast ensemble representations for abstract visual impressions. *Nature Communications, 7*, 13186. doi:10.1038/ncomms13186

Li, V., Herce Castañón, S., Solomon, J. A., Vandormael, H., & Summerfield, C. (2017). Robust averaging protects decisions from noise in neural computations. *PLOS Computational Biology, 13*, e1005723. doi:10.1371/journal.pcbi.1005723

Mareschal, I., Morgan, M. J., & Solomon, J. A. (2010). Attentional modulation of crowding. *Vision Research, 50*, 805–809.

Miller, A. L., & Sheldon, R. (1969). Magnitude estimation of average length and average inclination. *Journal of Experimental Psychology, 81*, 16–21.

Morgan, M. J., Hole, G. J., & Glennerster, A. (1990). Biases and sensitivities in geometrical illusions. *Vision Research, 30*, 1793–1810.

Myczek, K., & Simons, D. J. (2008). Better than average: alternatives to statistical summary representations for rapid judgments of average size. *Perception & Psychophysics, 70*, 772–788.

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature, 4,* 739–744.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437–442.

R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Raidvee, A., & Fougnie, D. (2017, August). Odd man out in perceptual averaging: How do outliers influence judgments? Poster session presented at the European Conference on Visual Perception 2017, Berlin, Germany. Retrieved January 6, 2019 from https://journals.sagepub.com/page/pec/collections/ecvp-abstracts/index/ecvp-2017

Spencer, J. (1961). Estimating averages. *Ergonomics, 4*, 317–328.

Spitzer, B., Waschke, L., & Summerfield, C. (2017). Selective overweighting of larger magnitudes during noisy numerical comparison. *Nature Human Behaviour, 1*, 1–8.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*, 207–232.

van den Berg, R., & Ma, W. J. (2012). Robust averaging during perceptual judgment is not optimal. *Proceedings of the National Academy of Sciences of the United States of America, 109*(13), E736.

Yeshurun, Y., & Carrasco, M. (1998). Attention improves or impairs visual performance by enhancing spatial resolution. *Nature, 396*, 72–75.

9 WHEN SUBJECTIVE IS THE MOST OBJECTIVE: A COMPARISON OF DIFFERENT FATIGUE MEASURES IN THE VISUAL MISMATCH NEGATIVITY (vMMN) TASK

KAIRI KREEGIPUU AND NELE PÕLDVER

We studied how fatigue induced by a time-on-task procedure lasting up to 90 minutes is related to different subjective and objective fatigue measures. The fatigue-inducing task consisted of a series of reaction time experiments within the visual mismatch negativity (vMMN) paradigm, where participants had to detect the visual motion of a grating in the central visual field and, depending on condition, either ignore or attend to visual motions in the background (see Kuldkepp, Kreegipuu, Raidvee, Näätänen, & Allik, 2013). Fatigue measures were subjective reports (Borg's Category-Ratio Scale, BCR-10), critical flicker fusion frequency (CFFF), and several eye-blink parameters (count, interval, duration, amplitude). Subjective reports of fatigue, CFFF, and the number and interval of blinks indicated that fatigue was increased during the experiment. Different fatigue measures did not significantly correlate, indicating that they probably represent different aspects of fatigue. Subjective reports of fatigue indicated the biggest effect size related to the dynamics of fatigue (i.e., feeling of tiredness preceding the decline in performance).

Introduction

Fatigue is highly common in contemporary society. It is one of the most frequent complaints in primary medical care (e.g., Bültmann, Kant, Kasl, Beurskens, & van den Brandt, 2002; DeLuca, 2005a) and is reported as being felt almost always or quite often

We thank our mentor Jüri Allik, whose experience in studying eye movements reaches back to 1970s, for inspiring us to look at eye-blink data. This research was supported by the Estonian Science Foundation (grant #8332), the Estonian Ministry of Education and Research (IUT02-13 and SF0180029s08), and the European Social Fund (Primus grant #3-8.2/60).

E-mail: kairi.kreegipuu@ut.ee and nele.poldver@ut.ee

by 42.7% of adult men and 48.5% of adult women in 2016 in Estonia[1]. Among other detrimental effects of fatigue, it has been estimated that 10–15% of all severe traffic accidents have been related to fatigue[2]. According to the prevalence of fatigue, this may even be a conservative estimate.

Fatigue is something that is much easier to experience than to define. It was Emil Kraepelin who observed in 1897 that a subjective feeling of tiredness (*Gefühl der Müdigkeit*) might not necessarily be reflected in objective indicators of fatigue (*Ermüdung*) (Kraepelin, 1897). There are more studies showing the independence of objective (i.e., indicated by a decrease in performance or established fatigue measures) and subjective (i.e., self-report) measures of fatigue, especially in clinical populations, than reports of a significant positive relationship between the two types of measures (DeLuca, 2005a). DeLuca (2005a) considers the lack of rigid correlation understandable because the feeling of tiredness may be compensated for by greater effort to keep the outcome of any performance unchanged. Only objective fatigue (i.e., real decrease in performance), viewed as a state of an organism's muscles, viscera, or central nervous system, which, as a result of physical activity and/or mental processing, or in the absence of sufficient rest, results in insufficient cellular capacity or system-wide energy to maintain the original level of activity and/or processing by using normal resources (Job & Dalziel, 2001) (i.e., *Ermüdung* in Kraepelin's terminology) is really dangerous to an organism. However, the subjective feeling of fatigue is an important warning sign that should not be overlooked, and needs properly reacting to, for example with rest or sleep.

Subjective and Objective Measures of Fatigue

There are many ways to measure fatigue and, besides the theoretical interest, it is also a practical question to determine which fatigue indicators are more sensitive to fatigue and which are less. The easiest way is to ask for a subjective rating of perceived fatigue (i.e., to use direct measures of fatigue). There are many standardized scales, for example, the Visual Analogue Scale (VAS; Monk, 1989), Borg's perceived intensity scales (The Borg Rating of Perceived Exertion, RPE, for perceived effort, and Borg's Category-Ratio scale, BCR-10, for fatigue; Borg, 1998), the Fatigue Assessment Scale (FAS; Michielsen, De Vries, & Van Heck, 2003), and the Multidimensional Fatigue Inventory (MFI-20; Smets, Garssen, Bonke, & De Haes, 1995; describing general fatigue, physical fatigue, reduced motivation, reduced activity, and mental fatigue). The Borg scales and VAS are quick one-item instruments to assess fatigue at any point in time.

One of the frequently used "objective" or non-self-reported fatigue measures is the critical flicker fusion frequency (CFFF; Curran, Hindmarch, Wattis, & Shillingford, 1990; Simonson & Brožek, 1952), describing the temporal resolution of the visual

[1] https://intra.tai.ee//images/prints/documents/149069399613_Eesti_taiskasvanud_rahvastiku_tervisekaitumise_uuring_2016.pdf (p. 124)

[2] https://www.swov.nl/en/facts-figures/factsheet/fatigue-traffic-causes-and-effects

system. There is a great amount of research suggesting that a rise in the CFFF threshold could be an indicator of central nervous system (CNS) activation and cortical arousal, whereas a fall could be associated with CNS fatigue phenomenon and a reduction in the efficiency of the system in processing information (Godefroy, Rousseu, Vercruyssen, Cremieux, & Brisswalter, 2002; Grego et al., 2005).

Eye blinks have also been shown to be valid and promising externally observable and "objective" indicators of fatigue (Caffier, Erdmann, & Ullsperger, 2003; Schleicher, Galley, Briest, & Galley, 2008), showing a higher rate of occurrence, longer duration and lower amplitude with increasing fatigue.

If several measures of fatigue are valid and reliable in indicating fatigue, they should correlate with each other. DeLuca recently (2005b) proposed using the degree of correlation between subjective and objective measures of fatigue to discriminate between primary and secondary fatigue. When the correlation is high, research is probably dealing with primary fatigue, which is caused by the primary neural mechanisms underlying fatigue (i.e., depletion of resources in the CNS or nerve damage). When the correlation is low, the relationship is contaminated by factors that selectively exacerbate fatigue effects, such as certain medications or a bad nights' sleep, resulting in so-called secondary fatigue.

Aims of the Present Study

The main aim of this study is to compare the different (subjective and objective) measures of fatigue and to determine which has the best validity in indicating fatigue in pre-attentive processing of visual stimuli (visual mismatch negativity, vMMN). Fatigue is induced by a time-on-task procedure (during a reaction-time task in the vMMN experiment), as it has been demonstrated that cognitive tasks lasting for about 2–3 hours cause fatigue (e.g., Boksem, Meijman, & Lorist, 2005; Lorist et al., 2000). Actually, it has been known for more than a century already that two hours or less of intensive mental work at a maximum efficiency lowers performance, but usually by not more than 10% (Thorndike, 1914). The MMN, discovered by Risto Näätänen and colleagues in 1978 (Näätänen, Gaillard, & Mäntysalo, 1978), is the characteristic relative negativity of an event related potential (ERP) curve representing differential processing of rare stimuli in comparison to the processing of frequent stimuli. Such processing difference, reflecting a general change detection ability of the brain, has been established in auditory, visual, tactile/somatosensory, and olfactory modalities (see Näätänen, Paavilainen, Rinne, & Alho, 2007, for a review). Recent research shows that vMMN is an emerging state marker (Kremláček et al., 2016) because it is disturbed in most neuropsychological diseases and states. Here, we take the results of our research on vMMN to motion direction changes (Kuldkepp, Kreegipuu, Raidvee, Näätänen, & Allik, 2013), and take a closer look into whether there is a relationship between vMMN parameters (i.e., latency or amplitude) and different fatigue measures indicating the alertness of the organism. For example, the processing capability of

visual motion and changes in it should correlate with the CFFF measurement, since both reflect the ability to process temporal variation. At the same time, both centrally originating indexes, CFFF and vMMN, have been shown to change in Alzheimer's disease (e.g., Curran & Wattis, 1998; Stothart, Kazanina, Näätänen, Haworth, & Tales, 2015; Tales & Butler, 2006), and have been related to glutamate (Näätänen et al., 2007; Rönnbäck & Hansson, 2004), indicating a possible common origin and vulnerability to fatigue.

How (v)MMN relates to fatigue, which is largely a reversible phenomenon, is not yet clear, as there are only a few studies on the topic. It has been shown that mental fatigue (i.e., generated by mental effort) is related to a decrease in auditory MMN amplitudes at fronto-central electrodes (Yang, Xiao, Liu, Wu, & Miao, 2013), or at temporal but not frontal electrodes (Wanyan, Zhuang, Lin, Xiao, & Song, 2018), and to a decrease in vMMN amplitudes in occipital and fronto-central electrodes (Li, Song, & Miao, 2018). Based on the findings of these studies, we approach the relationship between pre-attentive processing of visual stimuli (vMMN) and different fatigue measures with the following three hypotheses.

Hypothesis 1 (H1). The vMMN procedure (with a task to react to visual motion onsets) will induce fatigue that occurs at higher levels when measured by subjective reports than when measured by objective performance parameters (CFFF and eye blinks).

Hypothesis 2 (H2). All measures of fatigue (i.e., subjective ratings, CFFF, and eye blink parameters) are significantly correlated to each other.

Hypothesis 3 (H3). All fatigue measures correlate with pre-attentive information processing, specifically with vMMN amplitude, as this has been shown to be a state indicator.

H2 states that all the indicators used in our study measure something common, presumably fatigue. During the vMMN experiment, participants were presumed to get tired, and this state of tiredness is presumed to be accessible to participants' introspection (H1). However, we assume that the change in fatigue estimated by subjective ratings is bigger than fatigue measured by CFFF or blinks (H1), because the feeling of tiredness is considered to precede the objective decline in performance (DeLuca, 2005a). The pattern of correlations between the vMMN and fatigue measures (H3) allows the interpreting of probable fatigue dynamics (i.e., what fatigue measures relate more to the vMMN amplitude, and whether these correlations emerge at earlier or later time intervals).

Method

The detailed results of the vMMN experiment have been previously reported (Kuldkepp et al., 2013). Here, we look at the reported vMMN results in relation to additional and previously not published data on fatigue that were collected before and after the

experimental vMMN sessions lasting approximately 1.5 hours (with breaks). The vMMN experiment consisted of reaction time tasks where participants had to detect the visual motion of a grating in the central visual field and, depending on condition, either to ignore or attend to visual motions in the background (same as in Kuldkepp et al., 2013).

Participants

The results of the same volunteer subject group ($N = 49$) of healthy adults as in the Kuldkepp et al. (2013) study are reported here. Due to excessive artefacts or technical problems with the recordings in either the vMMN EEG experiment or resting state EEG measurement, the data of 6 subjects were excluded from the current analyses and thus the final sample consisted of 43 participants (67% females, mean age 21.3 years ($SD = 2.3$), range 19–27 years). All subjects reported to have normal or corrected-to-normal vision. They signed a written consent form and the study was approved by the Research Ethics Committee of the University of Tartu.

Data Collection and Measures

Electroencephalography (EEG) measurements. The subjects sat 90 cm from the monitor screen in a semi-darkened electrically shielded room. EEG was recorded with a BioSemi Active Two system (BioSemi, Amsterdam, Netherlands) using 32 active electrodes (placement based on the international 10/20 system; Jasper, 1958). Reference electrodes were placed on ear lobes. To register blinks and eye movements, a vertical electrooculogram was recorded with electrodes below and above the right eye and a horizontal electrooculogram with electrodes at the right and left outer canthi of the eyes. Online recording was done in DC mode with a 1024 Hz sample rate and a 0.16–100 Hz band-pass filter. Offline data analyses were performed using Brain Vision Analyzer 1.05 (Brain Products GmbH, Munich, Germany).

Eye-blink data collection and analysis. Eye-blink data were extracted from the resting state EEG recordings done before (pre) and after (post) the main vMMN experiment. Resting state EEG measurement protocol included both eyes-closed and eyes-open conditions (2 minutes each), but only the latter was used for the current analyses. During the eyes-open condition, the subjects were instructed to remain relaxed, avoid excessive body and eye movements and fixate on a black cross in the middle of a grey screen. Data were filtered offline from 1 to 13 Hz (24 dB/octave) using Butterworth zero phase filters. The start and end of eye-blinks were detected semi-automatically using a built-in Gratton and Coles algorithm (Gratton, Coles, & Donchin, 1983) to form blink segments. The data were baseline-corrected (from -250 ms before the start of the blink), and segment duration, from which the blink peak amplitude was detected, was 700 ms. The following blink parameters were exported and/or calculated for further analyses: blink duration, blink amplitude, number of blinks, and the time interval between two successive blinks. Due to artefacts or

technical reasons, the data from 6 subjects were not available for extraction, yielding 37 participants' data for blink parameters.

vMMN data collection and analysis. A detailed experimental procedure of the vMMN data collection and an overview of offline data analyses steps are reported in Kuldkepp et al. (2013). We registered vMMN to motion direction changes (direction change in 15% of cases acting as a deviant in an oddball paradigm) of a horizontally moving sine grating in the periphery of the computer screen. The subjects were instructed to perform a primary motion detection task in the center of the screen by pressing the button as soon as they saw the area moving, while the vMMN-eliciting stimulus presentation was at the periphery. The vMMN experiment consisted of different conditions that were presented randomly to the subjects. The experimental design was always the same; the only thing that varied was the task that was given to the subjects. Depending on the condition, the subjects either had to focus solely on the primary task (the Ignore condition, where they had to ignore the periphery and detect the motion onset of the central area), or focus on both and decide whether the two areas were moving in the same or opposite direction (the Attend condition). The rationale behind that division in the original report was to test whether vMMN is independent of attention (see Kuldkepp et al., 2013 for the results). vMMN mean amplitude was calculated in fixed 20-ms time windows from 60 to 400 ms. Due to the experimental design that kept Attend and Ignore conditions randomly intermittent, a comparison of vMMN parameters to find the within-session fatigue-related decrease in pre-attentive processing cannot be reliably performed.

Subjective fatigue. Subjective fatigue was measured before (pre) and after (post) the vMMN (EEG) experiment using a nonlinear BCR-10 scale for fatigue (Borg, 1998), ranging from 0 to 11 (from "nothing at all" to "extremely strong/maximal," respectively). The subjects had to indicate how tired they felt at the moment by writing down the corresponding number.

CFFF measurements. CFFF was measured with monocular vision using a subject's predefined dominant eye. Measurements were performed with a special apparatus to assess temporal resolution. The light stimulus was presented at a distance of 33 cm from the subject's eye through a dark non-reflective metal tube with a diameter of 2.7 cm, firmly attached to the light emitting diode. The subject was standing and looking into the tube from above. The opening of the tube was as close as possible to the eye, letting no outside light reach the eye. The frequency of the light emitting diode ranged from 14.2 to 48.9 Hz (being "flickering" and "steady", respectively) and could be increased or decreased manually in 0.1 Hz steps via a regulator button. Three ascending (flicker to steady) and three descending (steady to flicker) frequency values (presented alternately) were registered for each subject both before (pre) and after (post) the vMMN (EEG) experiment. The means of the three frequencies at which the sensation of flicker is replaced by perceived fusion and vice versa were calculated for pre and post measurements.

Results

A test for normality indicated that measures of pre-test subjective fatigue, post-test blink duration, and the average between-blink interval (in both sessions) did not follow normal distribution. Thus, in order to compare them, a non-parametric Wilcoxon Matched Pairs Test was used. For the other measures (CFFF, blink count, and amplitude), a parametric dependent samples *t*-test was used. First, we compared pre- and post-experiment fatigue measurements (H1). Fatigue during the experiment was indicated by subjective ratings, CFFF, number of blinks, and blink duration (see Table 9.1).

Table 9.1 Mean values (with *SD*) and differences (*t*- or *Z*-scores) for pre- and post-experiment fatigue measures.

Measure	Pre-test *M* (*SD*)	Post-test *M* (*SD*)	*N*	*t*/*Z*¤	*p*	Cohen's *d*[3]
Subjective fatigue on BCR-10 (0–11)	1.6 (1.2)	3.1 (1.4)	43	5.37¤	<.001**	1.34
CFFF (Hz)	35.6 (3.4)	34.7 (3.4)	43	3.72	<.001**	0.54
Blink count	20.3 (11.7)	25.5 (17.0)	37	-2.71	.010*	0.45
Blink duration (ms)	356.4 (46.9)	359.0 (56.3)	37	0.10¤	.922	0.06
Blink interval (ms)	6246.9 (3504.7)	5154.5 (4040.6)	36	2.53¤	.001*	0.31
Blink amplitude (μV)	191.5 (72.6)	196.3 (61.8)	37	-0.48	.634	0.08

Note. BCR-10 = Borg's Category-Ratio scale for fatigue; CFFF = critical flicker fusion frequency; depending on parametric or non-parametric (¤) comparison, *t*- or *Z*- statistic is used, respectively (see text for explanation).

**$p < .01$, *$p < .05$

Together with a significant difference between pre- and post-test values of measures and Cohen's *d*, our data indicate that, in our vMMN task—which we assumed to be fatigue-inducing—subjective reports of fatigue, CFFF, number of blinks, and blink interval are the best indicators of emerging fatigue, subjective reports being the most sensitive (as reflected in the largest Cohen's *d* value).

Second, we looked at the correlations between different fatigue measures (H2). All measures of fatigue showed considerable stability and surprising independence from each other, as can be seen in Table 9.2. The highest stability was shown by the CFFF ($r = .88$, Table 9.2), which differed significantly from test-retest correlations of subjective fatigue ($r = .64$, $p = .007$) and blink interval ($r = .58$, $p = .003$). Such stability of CFFF means that, although participants showed a decreased ability to discriminate flickering lights from a steady light in the post-test as compared to the pre-test (Table 9.1), they tended to change in ability similarly, irrespective of their initial value of CFFF.

[3] https://memory.psych.mun.ca/models/stats/eff ect_size.shtml

Only a few measures of fatigue showed significant correlations with each other. For example, both pre- and post-experiment subjective fatigue was significantly negatively correlated with post-experiment blink amplitude (i.e., participants who reported being more tired both before and after the experiment had blinks with smaller amplitudes after the experiment). Some blink parameters were significantly correlated with one another, but, for example, for blink interval and number of blinks, this was highly expected (because fewer longer-interval blinks and more shorter-interval blinks fit into two minutes). In addition, blink duration and amplitude showed some co-variability (i.e., significant correlations, see Table 9.2).[4]

Table 9.2 Pearson or Spearman correlations between different pre- and post-test fatigue measures.

							Blink					
	BCR-10*		CFFF		Count		Duration*		Interval*		Amplitude	
	Pre (1)	Post (2)	Pre (3)	Post (4)	Pre (5)	Post (6)	Pre (7)	Post (8)	Pre (9)	Post (10)	Pre (11)	Post (12)
1		**.64**	-.08	-.07	.19	.18	-.09	-.02	-.12	-.12	-.18	*-.33*
2			-.01	-.03	.10	.07	-.19	-.19	-.03	-.01	-.07	*-.35*
3				**.88**	-.03	-.05	.20	.09	.12	.09	.12	.16
4					-.18	-.12	.11	.13	.25	.11	-.01	.21
5						**.73**	.13	.23	**-.96**	**-.64**	.09	.04
6							.06	.17	**-.66**	**-.83**	.17	.04
7								**.63**	-.01	-.17	.46	*.42*
8									-.10	-.28	.23	.50
9										**.58**	-.05	.01
10											-.09	-.06
11												**.61**
12												

Note. * indicates Spearman correlations due to violations in normality. If not stated otherwise, Pearson correlations are used. BCR-10 = Borg's Category-Ratio scale for fatigue; CFFF = critical flicker fusion frequency. Correlations with *p*-values smaller than (a) .001 are shown in bold face, (b) .01 are underlined, and (c) .05 are in italics.

Next, we analyzed exploratively the correlations between vMMN amplitudes and different fatigue measures (H3, see Figure 9.1). According to our previous study, reliable vMMN was detected in an early time range (100–175 ms) and after 250 ms (see Table 1 in Kuldkepp et al., 2013). Thus, only these intervals are worth analyzing further. Roughly, only a correlation larger than .29-.30 is significant at $p < .05$ for a

[4] This general pattern of relative independence was also confirmed by exploratory factor analysis. In extracting principal components, these fatigue measures did not converge into one factor by any simple criterion (e.g., scree-test) and the optimal number of factors remained 3 or more, irrespective of input (pre- and post-test together or separately).

sample size of 37–43.[5] However, the average correlation between vMMN amplitudes in 11 relevant 20-ms intervals with 2 conditions and 4 electrodes (altogether 88 comparisons) remained below |.18| for all pre-test and post-test fatigue measures. Generally, in an ANOVA for these correlations (over all electrodes and comparisons), electrode Pz showed the highest correlations [$F(3, 132) = 5.73$, $p < .001$]. Even for Pz, only a few correlations were high enough to be considered further: these were average correlations within these intervals for vMMN amplitude and subjective fatigue in pretest, blink duration in post-test, and average blink amplitude in post-test conditions, which were $r = .23$ (95% CI .19 to .29), $r = -.23$ (95% CI -.30 to -.16), and $r = .27$ (95% CI -.34 to -.20), respectively. When the duration criteria of the significant correlation at $p < .05$ was set to two consequent intervals (i.e., at least 40 ms), there were three such correlations for Pz. In the Attend condition, vMMN amplitude up to 260 ms was related to pre-test subjective fatigue, and post-test blink amplitude. Post-test blink amplitude was related to vMMN amplitude in the later time range (360–400 ms) as well. In the Ignore condition, post-test blink amplitude was related to most of the vMMN amplitudes registered in Pz (see Figure 9.1).

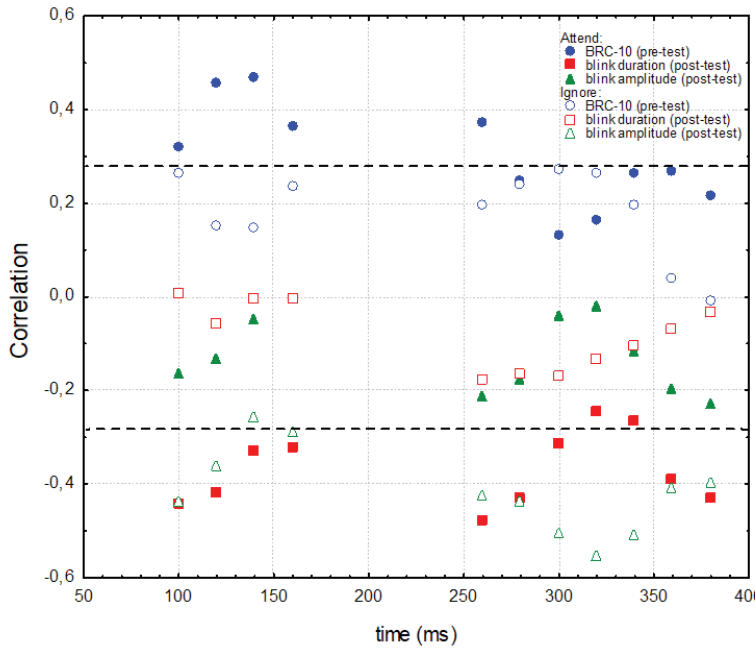


Figure 9.1 Spearman correlations between fatigue measures and vMMN amplitude (electrode Pz) in Attend (filled shapes) and Ignore (empty shapes) conditions. vMMN correlations with BCR-10 pre-test score are marked with circles, post-test blink duration with squares, and post-test blink amplitude with triangles. Only intervals with significant vMMN (see Kuldkepp et al., 2013) and correlations exceeding a significance level (.29-.30 at $p < .05$, marked by broken lines) in any condition are presented. BCR-10 = Borg's Category-Ratio scale for fatigue.

[5] https://www.york.ac.uk/depts/maths/tables/spearman.pdf

In interpreting these correlations, it is important to remember that vMMN is an ERP curve with negative amplitude. Thus, the more positive the vMMN amplitude is, the less vMMN there actually is (the negativity of the vMMN curve is smaller in amplitude), referring to more fatigue. This is exactly the pattern Figure 9.1 represents for subjective fatigue: less vMMN (i.e., more positive amplitude) is related to higher levels of subjective pre-test fatigue (circles, positive correlation). For eye blinks (duration and amplitude), the pattern is more inconsistent, which can be explained by the fact that these indicators did not seem to be affected by the increased level of fatigue during the vMMN task (see Cohen's *d* in Table 9.1). Post-test blink amplitude seems to relate negatively to vMMN amplitude in the Ignore condition: smaller vMMN (i.e., more positive amplitude) indicated smaller blink amplitude, a relationship possibly mediated by fatigue. For blink duration, the negative correlation to vMMN amplitude in the Attend condition is surprising, but this may be related to the requirement to deliberately move the eyes to take into account the events in the background in this condition.

Discussion

The analysis of the associations between fatigue and vMMN shows three important things. First, the experimental session in a dimly lit room lasting approximately 1.5 hours (pre- and post-experiment measurements and resting pauses included) creates some fatigue (see Table 9.1). In this regard, we can think of the vMMN main experiment itself as a mental fatigue-inducing task (cf. Boksem et al., 2005; Lorist et al., 2000, Thurstone, 1914). Different measures of fatigue do differ in their sensitivity to detect or measure fatigue: subjective reports of fatigue seem to precede the objective decline in performance. This generalizes presumably to real life, too: people report fatigue considerably more often than they actually fail in their performance (cf. DeLuca, 2005a). Thus, H1 was confirmed.

Second, the fatigue measures used had relatively high test-test reliability (Table 9.2), which was higher for measures reacting more sensitively to fatigue (subjective fatigue and CFFF, indicated by Cohen's *d*s in Table 9.1). This may seem a bit paradoxical, but it just means that the more reliable the fatigue measures are, the more similar most participants' reactions to fatigue are. For example, if participants had reported more fatigue in the post-test than in the pre-test, the high reliability of a measure would mean that all participants tended to increase in this subjective dimension similarly, and their ranking did not change much.

In addition to different fatigue curves (i.e., the speed with which fatigue is elicited/risen), it may be true that different fatigue measures reflect different sub-types of fatigue (primary, secondary, mental, motivational, etc.), or fatigue in different parts/functions of the organism. This is supported by the relative independence of the observed fatigue

measures (Table 9.2), and thus, H2 was not supported. For example, self-reported or subjective fatigue is more likely a reflection of secondary fatigue, and CFFF of primary fatigue, relying possibly on different types of cells in the visual system (Wells, Bernstein, Scott, Bennett, & Mendelson, 2001). The fact that different measures of a construct do not correlate with each other is not restricted to fatigue. One of the well-known examples is the construct of impulsivity (e.g., Dougherty, Mathias, Marsh, & Jagar, 2005; Havik et al., 2012; Reynolds, Ortengren, Richards, & de Wit, 2006), which typically shows low correlations between impulsivity measured by questionnaires and behavioral laboratory tests. Of course, it is also possible that our measures differ on how much they do reflect fatigue, but this aspect remains to be analyzed in later studies.

Third, vMMN as a sensitive indicator of the state of the whole organism (H3) was also not supported by the findings of this study (Figure 9.1). In this study, though, the random presentation of vMMN conditions (Ignore and Attend) during the EEG experiment did not allow specifically for the tracking of the possible presentation order effects of the two conditions, which might also have contaminated the pattern of correlations. The Attend condition demanded more attentional resources from the participants (as they had to divide their attention between two different moving areas—center and periphery), so we can assume that this condition generates more fatigue than the Ignore condition with a simple reaction-time task. The two conditions presented different demands on eye movements, as, in the Ignore condition, the participants had to look only at the center of the screen, but in the Attend condition, they had to track movement in both the center and the periphery. These differences in the setup of the experimental conditions likely contributed to vMMN and its correlations with eye-blink parameters. Self-reported fatigue, CFFF, and blinks may all be related to the automatic visual processing of stimuli, but this relationship needs to be studied with experiments specifically targeting this question.

Limitations and Conclusions

Altogether, the study shows that subjective reports of fatigue (BCR-10) are a reasonably good indicator of fatigue as this stood out in all three tests used. It showed the highest effect size when representing change in the state of participants (Table 9.1), had reasonable test-retest reliability (r = .64, Table 9.2), and was related with remarkable stability to preattentive processing of stimuli (vMMN amplitude, Figure 9.1).

The current report has some limitations that should be considered, and taken into account when planning future studies. First, the focus of the original data collection was on the vMMN as the main experiment, which means that the timing of the pre- and post-measurements concerning fatigue slightly varied across subjects, depending on their individual need to take breaks between different series of the vMMN experiment. In addition, the vMMN experiment conditions (Ignore and Attend, reported in Kuldkepp et al., 2013, as well as two other unreported conditions) were presented to subjects in random order. To pinpoint changes in different fatigue measurements more precisely,

experimental procedures should be homogenous across the subject group. Currently, our study remains explorative with respect to fatigue measures.

Second, the CFFF thresholds were obtained with monocular vision. It has been shown, however, that CFFF values are higher when measured with binocular rather than monocular vision (Ali & Amir, 1991). Higher CFFF levels might be more vulnerable to fatigue than lower levels.

Third, there remains the question of when should vMMN be recorded to target fatigue effects in the vMMN amplitude or latency parameters. In our study, fatigue measurements were recorded before and after the vMMN experiment, meaning that the vMMN experiment itself could be considered as a fatigue-inducing task. A second and more precise option is to record vMMN and fatigue measures both before and after a separate fatigue-inducing task, as was done by Li and colleagues (2016), for instance.

To the best of our knowledge, correlations between vMMN and fatigue measurements had not been reported before. The study by Li and colleagues (2016) presented the results of subjective fatigue and vMMN measurements before and after mental fatigue manipulation, but they did not report the relationship between the measurements. Thus, we believe that the current report, despite of its limitations, adds to this topic and suggests a necessary direction in vMMN research. The knowledge of the possible different measurements reflecting the same underlying state of the CNS and information processing capacity has direct practical implications in being able to choose between different methods (including choosing less time- or money-consuming or less invasive options) and save time in collecting data with only one method instead of many.

References

Ali, M. R., & Amir, T. (1991). Critical flicker frequency under monocular and binocular conditions. *Perceptual and Motor Skills*, *72*, 383–386.

Boksem, M. A., Meijman, T. F., & Lorist, M. M. (2005). Effects of mental fatigue on attention: An ERP study. *Cognitive Brain Research*, *25*, 107–116.

Borg, G. (1998). *Borg's Perceived Exertion and Pain Scales*. Champaign, IL: Human Kinetics.

Bültmann, U., Kant, I., Kasl, S. V., Beurskens, A. J., & van den Brandt, P. A. (2002). Fatigue and psychological distress in the working population: Psychometrics, prevalence, and correlates. *Journal of Psychosomatic Research*, *52*, 445–452.

Caffier, P. P., Erdmann, U., & Ullsperger, P. (2003). Experimental evaluation of eye-blink parameters as a drowsiness measure. *European Journal of Applied Physiology, 89*, 319–325.

Curran, S., Hindmarch, I., Wattis, J. P., & Shillingford, C. (1990). Critical Flicker Fusion in normal elderly subjects: A cross-sectional community study. *Current Psychology, 9*, 25–34.

Curran, S., & Wattis, J. P. (1998). Critical flicker fusion threshold: A useful research tool in patients with Alzheimer's disease. *Human Psychopharmacology: Clinical and Experimental*, *13*, 337–355.

DeLuca, J. (2005a). Fatigue, cognition and mental effort. In J. DeLuca (Ed.), *Fatigue as a window to the brain* (pp. 37–57). Cambridge, MA: The MIT Press.

DeLuca, J. (2005b). Fatigue: Its definition, its study and its future. In J. DeLuca (Ed.), *Fatigue as a window to the brain* (pp. 319–325). Cambridge, MA: The MIT Press.

Dougherty, D. M., Mathias, C. W., Marsh, D. M., & Jagar, A. A. (2005). Laboratory behavioral measures of impulsivity. *Behavior Research Methods*, *37*, 82–90.

Godefroy, D., Rousseu, C., Vercruyssen, F., Cremieux, J., & Brisswalter, J. (2002). Influence of physical exercise on perceptual response in aerobically trained subjects. *Perceptual and Motor Skills*, *94*, 68–70.

Gratton, G., Coles, M. G. H., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*, 468–484.

Grego, F., Vallier, J. M., Collardeau, M., Rousseu, C., Cremieux, J., & Brisswalter, J. (2005). Influence of exercise duration and hydration status on cognitive function during prolonged cycling exercise. *International Journal of Sports Medicine, 26,* 27–33.

Havik, M., Jakobson, A., Tamm, M., Paaver, M., Konstabel, K., Uusberg, A., Allik, J., Ööpik, V. & Kreegipuu, K. (2012). Links between self-reported and laboratory behavioral impulsivity. *Scandinavian Journal of Psychology* 53, 216–223.

Jasper, H. H. (1958). The ten-twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology, 10*, 370–375.

Job, R. F. S., & Dalziel, J. (2001). Defining fatigue as a condition of the organism and distinguishing it from habituation, adaptation, and boredom. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, workload, and fatigue* (pp. 466–475). Mahwah, NJ: Lawrence Erlbaum Associates.

Kraepelin, E. (1897). *Zur Überbürdungsfrage.* Jena: Verlag von Gustav Fischer.

Kremláček, J., Kreegipuu, K., Tales, A., Astikainen, P., Põldver, N., Näätänen, R., & Stephanics, G. (2016). Visual mismatch negativity (vMMN): A review and meta-analysis of studies in psychiatric and neurological disorders. *Cortex, 80,* 76–112.

Kuldkepp, N., Kreegipuu, K., Raidvee, A., Näätänen, R., & Allik, J. (2013). Unattended and attended visual change detection of motion as indexed by event-related potentials and its behavioral correlates. *Frontiers in Human Neuroscience, 7, 476*. doi:10.3389/fnhum.2013.00476

Li, J., Song, G., & Miao, D. (2016). Effect of mental fatigue on nonattention: A visual mismatch negativity study. *NeuroReport*, *27*, 1323–1330.

Lorist, M. M., Klein, M., Nieuwenhuis, S., De Jong, R., Mulder, G., & Meijman, T. F. (2000). Mental fatigue and task control: planning and preparation. *Psychophysiology*, *37*, 614–625.

Michielsen, H. J., De Vries, J., & Van Heck, G. L. (2003). Psychometric qualities of a brief self-rated fatigue measure: The Fatigue Assessment Scale. *Journal of Psychosomatic Research*, *54*, 345–352.

Monk, T. H. (1989). A visual analogue scale technique to measure global vigor and affect. *Psychiatry Research*, 27, 89–99.

Näätänen, R., Gaillard, A. W., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica, 42*, 313–329.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology, 118*, 2544–2590.

Reynolds, B., Ortengren, A., Richards, J. B., & de Wit, H. (2006). Dimensions of impulsive behavior: Personality and behavioral measures. *Personality and Individual Differences*, *40*, 305–315.

Rönnbäck, L., & Hansson, E. (2004). On the potential role of glutamate transport in mental fatigue. *Journal of Neuroinflammation*, *1*, 22. doi:10.1186/1742–2094-1-22

Schleicher, R., Galley, N., Briest, S., & Galley, L. (2008). Blinks and saccades as indicators of fatigue in sleepiness warnings: Looking tired? *Ergonomics*, *51*, 982–1010.

Simonson, E., & Brožek, J. (1952). Flicker fusion frequency: Background and applications. *Psychologial Reviews, 32,* 349–378.

Smets, E. M. A., Garssen, B., Bonke, B., & De Haes, J. C. J. M. (1995). The multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. *Journal of Psychosomatic Research*, *39*, 315–325.

Stothart, G., Kazanina, N., Näätänen, R., Haworth, J., & Tales, A. (2015). Early visual evoked potentials and mismatch negativity in Alzheimer's disease and mild cognitive impairment. *Journal of Alzheimer's Disease*, *44*, 397–408.

Tales, A., & Butler, S. (2006). Visual mismatch negativity highlights abnormal preattentive visual processing in Alzheimer's disease. *NeuroReport*, *17*, 887–890.

Thorndike, E. L. (1914). *Educational psychology, Vol. 3. Mental work and fatigue and individual differences and their causes*. New York: Teachers College.

Yang, B., Xiao, W., Liu, X., Wu, S., & Miao, D. (2013). Mental fatigue impairs pre-attentive processing: AMMN study. *Neuroscience Letters*, *532*, 12–16.

Wanyan, X., Zhuang, D., Lin, Y., Xiao, X., & Song, J. W. (2018). Influence of mental workload on detecting information varieties revealed by mismatch negativity during flight simulation. *International Journal of Industrial Ergonomics*, *64*, 1–7.

Wells, E. F., Bernstein, G. M., Scott, B. W., Bennett, P. J., & Mendelson, J. R. (2001). Critical flicker frequency responses in visual cortex. *Experimental Brain Research*, *139*, 106–110.

10 CAN'T WAIT OR WANT IT NOW? IMPULSIVITY RELATES TO THE IMMEDIACY RATHER THAN THE DELAY SENSITIVITY ASPECT OF TEMPORAL DISCOUNTING

ANDERO UUSBERG, UKU VAINIK, AND KAIRI KREEGIPUU

Behavior in delay discounting experiments is less consistently related to trait impulsivity than the conceptual overlap between these constructs implies. We tested whether the behavior-trait correlation can be improved by decomposing discounting behavior into a pair of scores representing the overvaluation of immediate rewards (Immediacy Sensitivity) and undervaluation of delayed rewards (Delay Sensitivity). Fifty-six students made choices between imaginary monetary rewards with the sooner option being either available immediately (now-later condition) or delayed (later-later condition). The area over the discounting curve (AOC) in the later-later condition was used as a measure of Delay Sensitivity, while the difference between later-later and now-later AOCs represented Immediacy Sensitivity. Compared to estimates from traditional two-parameter discounting models, Immediacy Sensitivity was a more consistent predictor of self-reported impulsivity as well as behavioral impulsivity in the domain of alcohol consumption. We conclude that Immediacy Sensitivity may be closer to the core of trait impulsivity than traditional discounting rates or Delay Sensitivity.

Introduction

Life requires choosing between immediate temptations and delayed gratifications. People generally opt for the former, in effect discounting the subjective value of rewards as a function of their delay (Scheres, de Water, & Mies, 2013). This tendency varies systematically between individuals, suggesting that intertemporal decision-making may be an important behavioral reflection of trait impulsivity (Peters & Büchel, 2011). However, the observed correlations between discounting behavior and self-reported impulsivity have been inconsistent (Duckworth & Kern, 2011; Mahalingam, Stillwell, Kosinski, Rust, & Kogan, 2014; Sharma, Markon, & Clark, 2014). Possibly, this is because individual discounting differences arise from a mixture

We thank Annika Teder, Anu Ilves, Egert Valmra, Kadi Tulver, Kerttu Mäger, and Kristjan Mesipuu for their help in conducting the experiment and Kenn Konstabel for his insightful feedback on its interpretation. This study was supported by grants from the European Social Fund (Primus grant #3-8.2/60) and the Estonian Research Council (SF0180029s08; IUT02-13).

E-mail: andero.uusberg@ut.ee

of underlying mechanisms, only some of which are relevant for trait impulsivity. In particular, intertemporal decisions involve (a) computing the present subjective value of delayed rewards, and (b) controlling an additional drive towards immediate outcomes (Berns, Laibson, & Loewenstein, 2007; Metcalfe & Mischel, 1999; Peters & Büchel, 2011; Scheres et al., 2013). We propose a behavioral method for decomposing individual differences in discounting into two scores—Delay Sensitivity and Immediacy Sensitivity—reflecting these respective sources of variance. We ask which score better predicts self-reported impulsivity (Whiteside & Lynam, 2001) as well as behavioral impulsivity in the context of alcohol consumption. We also compare the validity of this approach to parameter estimates from hyperboloid (Rachlin, 2006) and quasi-hyperbolic (Laibson, 1997) discounting models.

This work was conducted during the graduate studies of the first author, which were inspired, enabled, and supervised by the reason for this book—Professor Jüri Allik. After having searched together for the EEG correlates of the Kinematic Energy Model of motion perception (Dzhafarov, Sekuler, & Allik, 1993), Andero and Jüri turned their attention to the psychological mechanisms of personality traits, or the mental processes that turn traits into behavior. Their aim was to help bridge an important gap in personality research. On the one hand, it had become undeniable that variance in the Big Five personality traits of Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness is highly heritable (Allik, 2002). On the other hand, it was, and largely still is, unclear what mechanisms mediate the correspondence between genotypic and phenotypic patterns of personality. One possibility is that genotypic patterns govern the operating parameters of certain brain systems and that these parameters, in turn, govern the phenotypic patterns of thought, feeling, and behavior captured by personality questionnaires (DeYoung, 2010). If this were true, then it should be possible to devise experiments that reveal the values of these key parameters. The study presented in this chapter is one attempt to do just that.

Even though this study is underpowered, we believe it showcases some of Jüri's key scientific dispositions. First, the study operates across several areas of analysis including personality, behavior, and neuroscience. This resembles, on a small scale, Jüri's exceedingly rare capacity to understand an impressive range of phenomena, from perception to emotion, and to work with an equally impressive range of tools developed in disciplines from genetics to human geography. Second, the temporal discounting literature our study relies on, shares Jüri's affection for explicit mathematical models of psychological processes. Further in line with Jüri's affections, we find that a simpler mathematical formulation works better than more complex alternatives. Finally, this study exemplifies a principle that Jüri has helped us understand—that, in science, data should ultimately serve ideas, and not the other way around.

Individual Differences in Delay Discounting

Personality traits such as conscientiousness and impulsivity capture individual differences in tendencies to pursue long-term over short-term goals. It is therefore reasonable to expect that individuals scoring high on conscientiousness or low on impulsivity would exhibit lower discounting of delayed rewards. Individual differences in discounting behavior indeed contain a relatively stable trait component (Peters & Büchel, 2011), as demonstrated by their reasonable stability across reward domains (Odum, 2011) and measurements (Beck & Triplett, 2009; Kirby, 2009). Discounting rates also correlate with impulsivity-related outcomes, such as substance abuse, gambling, and health behavior (MacKillop et al., 2011; Reynolds, 2006; Story, Vlaev, Seymour, Darzi, & Dolan, 2014). Intriguingly, however, correlations between discounting rates and self-reported impulsivity are relatively inconsistent. Next to observed correlations between discounting rates and self-reported measures of impulsiveness (Bobova, Finn, Rickert, & Lucas, 2009), extraversion, sensation seeking (Hirsh, Morisano, & Peterson, 2008; Ostaszewski, 1996), conscientiousness, and neuroticism (Manning et al., 2014), there are also failures to replicate such relationships (Reynolds, Ortengren, Richards, & de Wit, 2006; Swann, Bjork, Moeller, & Dougherty, 2002). As a result, meta-analytic summaries of the available evidence reveal only a weak relationship between discounting behavior and self-reported impulsivity (Duckworth & Kern, 2011; Mahalingam et al., 2014; Sharma et al., 2014).

What could explain the inconsistent relationships between behavioral delay discounting and self-reported impulsivity? One possibility is that discounting behavior is driven by more than one underlying process (e.g., Green & Myerson, 2013) and only some of these processes relate to trait impulsivity. This interpretation is supported by evidence for the involvement of a number of different brain networks in inter-temporal choices (Luhmann, 2009; Peters & Büchel, 2011; Scheres et al., 2013; Sellitto, Ciaramelli, & di Pellegrino, 2011). In particular, a distinction has often been made between networks involved in valuation, on the one hand, and those involved in motivational salience, on the other (Berns et al., 2007; Liu, Feng, Wang, & Li, 2012; Luhmann, 2009; McClure, Laibson, Loewenstein, & Cohen, 2004; Metcalfe & Mischel, 1999; Peters & Büchel, 2011). We relate this neural distinction to a conceptual distinction between two aspects of delay discounting: Delay Sensitivity reflecting individual differences in valuation and Immediacy Sensitivity reflecting individual differences in motivational salience.

The valuation component of inter-temporal choice first involves a domain-general midbrain network (Bartra, McGuire, & Kable, 2013; Levy & Glimcher, 2012) associated with representing the delayed and immediate rewards in the "common currency" of subjective value (Kable & Glimcher, 2007; Liu et al., 2012; Sellitto et al., 2011; Sripada, Gonzalez, Luan Phan, & Liberzon, 2011). Second, computing the present value of delayed rewards also activates brain areas involved in predicting and simulating future outcomes (Peters & Büchel, 2011). We consider these networks as a functionally

integrated unit of analysis whose variance between individuals underlies Delay Sensitivity. We hypothesize that high levels of Delay Sensitivity manifest in difficulties with representing the true present value of delayed rewards (Ballard & Knutson, 2009; Peters & Büchel, 2011; Sripada et al., 2011).

Another origin of impatience—Immediacy Sensitivity—is hypothesized to involve insufficient control of the motivational salience of immediate rewards (Berns et al., 2007; Peters & Büchel, 2011; Story et al., 2014). Immediate availability of rewards is believed to increase their motivational salience, which can impact choices over and above the present subjective value of alternatives computed by the valuation networks (Benhabib, Bisin, & Schotter, 2010; Luo, Ainslie, Giragosian, & Monterosso, 2009). This immediacy bias has been associated with a combination of overactive midbrain dopaminergic pathways and underactive prefrontal regulatory pathways (Essex, Clinton, Wonderley, & Zald, 2012; Figner et al., 2010; Liu et al., 2012). We therefore consider Immediacy Sensitivity a reflection of characteristically strong motivational salience or characteristically weak prefrontal control, or a combination of the two. We hypothesize that high levels of Immediacy Sensitivity manifest in difficulties in overcoming the immediacy bias.

Decomposing Discounting Behavior

In addition to the conceptual distinction between Delay and Immediacy Sensitivity, we propose a way to measure these traits behaviorally, without the help of neuroimaging. In a typical delay discounting task, participants make a series of choices between a Smaller Sooner Reward (SSR, e.g., 12€ available now) and a Larger Later Reward (LLR, e.g., 24€ available in 4 weeks). By systematically varying these values, an SSR can be found that is chosen equally often with a given LLR (i.e., an indifference point). The indifference point can be taken to represent the subjective present value of the delayed reward (Smith & Hantula, 2008). Plotting the subjective present values of the same reward for different delays reveals a discounting curve. The steepness of that curve, or the discounting rate, reflects the extent to which a given individual discounts the value of rewards that are delayed.

In order to disentangle Delay and Immediacy Sensitivity contributions to the discounting curve, we rely on an additional experimental condition with a front-end delay of 2 weeks added to both rewards (e.g., a now-later choice between "12€ now or 24€ in 4 weeks" becomes a later-later choice between "12€ in 2 weeks or 24€ in 6 weeks"; cf. Green, Myerson, & Macaux, 2005; Kirby & Herrnstein, 1995; Luhmann, 2013; McClure et al., 2004; Sripada et al., 2011). We assume that the discounting rate variance in the traditional now-later condition is an additive function of individual differences in computing the present value of delayed rewards (i.e., Delay Sensitivity), as well as in controlling the immediacy bias (i.e., Immediacy Sensitivity). The observed variance in the delayed condition, by contrast, should originate only from Delay Sensitivity, since, in the absence of an immediate option, the mechanisms involved in

the immediacy bias should remain relatively inactive. Scores reflecting different aspects of discounting behavior can therefore be obtained by (a) treating the discounting rate of the later-later condition as a Delay Sensitivity score, and (b) subtracting this score from the now-later discounting rate to yield an Immediacy Sensitivity score.

Aims of the Study

We report a small-scale test of the idea that consistent correlations between trait impulsivity and delay discounting behavior can be revealed by decomposing the latter into Delay Sensitivity and Immediacy Sensitivity scores. Specifically, we extracted these scores from a behavioral experiment and analyzed their correlations with self-reported and behavioral impulsivity in comparison to traditional model-based alternatives.

Self-reported impulsivity was assessed using the UPPS model, which differentiates 4 facets: Urgency; (lack of) Planning; (lack of) Perseverance, and Sensation Seeking (Whiteside & Lynam, 2001). As an instance of behavioral impulsivity, we focused on alcohol consumption, which is relatively prevalent among young adults (Kuntsche, Knibbe, Gmel, & Engels, 2006), and has been related to both self-reported impulsivity (Stautz & Cooper, 2013) and discounting rates (MacKillop et al., 2011; Rossow, 2008).

Immediacy and Delay Sensitivity can be considered alternatives to existing model-based methods for decomposing discounting variability into more than one meaningful parameter (Doyle, 2013; Franck, Koffarnus, House, & Bickel, 2015). We will therefore compare the approach proposed here to a pair of such models. First, a popular and successful (Green & Myerson, 2004; McKerchar et al., 2009; Takahashi, 2009) hyperboloid model will be investigated (Rachlin, 2006)[1]:

$$V = A/(1 + kD^s) \qquad \textit{(Equation 1).}$$

In this model, V is the subjective present value of reward A delayed by time D. Parameter k reflects the reduction of value induced by each unit of delay (i.e., the discounting rate), while parameter s has been associated with the logarithmic relationship between perceived and actual time (Zauberman, Kim, Malkoc, & Bettman, 2009).

The second model under consideration is the quasi-hyperbolic or beta-delta model (Berns et al., 2007; Laibson, 1997; Phelps & Pollak, 1968):

$$\begin{Bmatrix} if\ D = 0 & V = A \\ if\ D > 0 & V = A\beta\delta^D \end{Bmatrix} \qquad \textit{(Equation 2).}$$

This model can be viewed as an attempt to capture the variability we refer to as Delay and Immediacy Sensitivity in separate parameters of a single quasi-hyperbolic

[1] This particular form of the hyperboloid function provided a better fit for the present data than an alternative where the exponent s is applied to whole denominator (Myerson & Green, 1995).

function. The model assumes that, at zero delay, *V* equals *A*. At each subsequent delay, *V* is a function of an exponential discounting rate δ^D and a special penalty β induced by non-immediacy (i.e., the inverse of immediacy bias). Parameter estimates from these models will be used as benchmarks against which to test the newly defined Immediacy and Delay Sensitivity scores.

Methods

After removing 6 participants with aberrant discounting curves (in line with an algorithm proposed by Johnson & Bickel, 2008), the sample consisted of 56 healthy university students (mean age 22 years, *SD* = 3.70, range 19–41, 15 males). Personality and alcohol consumption instruments were administered 3 to 7 days before the experiment in an online environment. The delay discounting experiment was administered together with a stop-signal task and a time perception task during an electroencephalographic recording session in a quiet, dark room at 1 m distance from a 19-inch CTR monitor. The study was approved by the Ethics Review Committee of Human Research of the University of Tartu. The data were collected in 2010.

In the discounting experiment analyzed here, participants indicated their preferences in 180 hypothetical inter-temporal choices while imagining that the offers were real. On each trial, two choice options were presented on either side of the screen: the SSR on the left and the LLR on the right (e.g., "700 EEK now or 950 EEK in 2 weeks"). Pairs of consecutive trials were randomly selected from 18 blocks formed by combining 2 SSR delays (0 in the now-later condition; 2 weeks in the later-later condition), 3 relative LLR delays (2, 4, or 10 weeks from the SSR delay) and 3 LLR values (925, 950, or 975 EEK, equivalent to €59.10, €60.70, and €62.30 respectively). The SSR values were set adaptively to maximize difficult choices by continuously shortening the vector of possible indifference points (IDP) of a given block (Wittmann, Leland, & Paulus, 2007)[2].

[2] Before the first trial, the IDP must lie somewhere between 0 and LLR (e.g., 952 EEK). The adaptive algorithm divided this vector into three equal sections *a*, *b*, and *c* and presented the rounded cut-off points (e.g., 317 between *a* and *b*, 635 between *b* and *c*) as SSRs on two consecutive trials to identify whether the IDP falls within *a*, *b*, or *c*. If the subject preferred the delayed 952 to both the sooner 317 and sooner 635 EEK, their IDP should lie within *c*. If they preferred both 635 and 317 to the delayed 952 EEK, their IDP should lie in *a*. Finally, if they traded the delayed 952 for the sooner 635 but not for 317 EEK, their IDP should be within *b*. The identified section was then taken as the new vector of possible IDP values and divided again into three sections. After five iterations of this procedure, a final IDP was defined as the midpoint of the final vector. In the case of an inconsistent choice pattern, the previous pairing was presented again.

To calculate the Delay and Immediacy Sensitivity scores, data from the three LLR values were averaged. Areas over the discounting curves (AOC[3]) were estimated for now-later as well as later-later conditions as 1 minus area under the curve, assessed using the trapezoid method (Myerson, Green, & Warusawitharana, 2001). Delay Sensitivity was then defined as the AOC from the later-later condition and Immediacy Sensitivity as the now-later AOC minus the later-later AOC. For ease of interpretation, both sensitivities were converted to *z*-scores for subsequent analyses. The model parameters were estimated for IDPs from the now-later condition using the nonlinear least squares estimation function of the R statistical language (R Core Team, 2014)[4]. The mean model fit for the hyperboloid model was R^2 = .98 (ranging between .82 and .99), and for the quasi-hyperbolic model R^2 = .89 (ranging between .28 and .99).

Self-reported impulsivity was measured using relevant facets of the Five Factor Model of personality: Urgency with N5: Impulsiveness; (lack of) Planning with C6: Deliberation; (lack of) Perseverance with C5: Self-Discipline; and Sensation Seeking with E5: Sensation Seeking (Whiteside & Lynam, 2001). The facets scores were measured with the EE.PIP-NEO inventory (Mõttus, Pullmann, & Allik, 2006) and standardized according to age- and gender-specific normative data, published elsewhere (normative sample age range 19–29, n = 369 for females and n = 733 for males; Mõttus et al., 2006). The questionnaire was completed by 54 participants.

Alcohol consumption was assessed using the graduated frequency approach (Dawson & Room, 2000), whereby participants indicated the yearly frequency (almost every day; 3–4 times a week; 1–2 times a week; 2–3 times a month; once a month; 6–11 times a year; 1–5 times a year; not at all) of drinking certain types and amounts of alcohol, as well as encountering alcohol-related problems (i.e., accidents, fights, personal problems, professional problems, and health problems). Three variables were derived from these responses. Overall consumption was expressed as the number of liters of pure ethanol consumed over the last 12 months. Binge drinking was expressed as the proportion of total alcohol consumed while drinking more than 35 g of ethanol for women and 60 g for men in one session (Dawson & Room, 2000). The prevalence of drinking problems was converted *post-hoc* into a three-level factor: no drinking problems (n = 17); up to 3 drinking problems (n = 16); and more than 3 drinking problems (n = 14). Alcohol consumption data were available for 47 participants.

[3] Area over, rather than under, the curve was preferred for semantic consistency - higher values on AOC reflect higher levels of impulsivity.

[4] Fitting used the Gauss–Newton algorithm for Equation 1 and the adaptive nonlinear least-squares algorithm (Dennis, Gay, & Walsh, 1981) for Equation 2 in order to constrain the β parameter between 0 and 1 and δ below 1 (McClure et al., 2004). Starting values required by the algorithm were initially set at 0.5 for all parameters. If the algorithm failed to converge, starting values were iteratively increased (odd iterations) and decreased (even iterations) by 0.05 until a suitable fit was obtained.

Results

We begin by inspecting the discounting curves computed for the two conditions of this study. Figure 10.1 depicts the mean present values (i.e., indifference points) of 950 EEK as a function of relative delays in the now-later and later-later conditions. The discounting curves illustrate how the present value decreases with increasing delay in both conditions. Although now-later choices often produce steeper discounting than later-later choices (e.g., Green et al., 2005), in our sample, the mean areas over the discounting curves did not differ significantly between now-later (M = .61, SD = .17) and later-later (M = .62, SD = .17) conditions (Sign test p < .89).

The AOCs from now-later and later-later conditions were highly correlated (Spearman r = .88, p < .01), suggesting that behavior in these conditions is indeed determined by partially overlapping sources. By contrast, the Immediacy and Delay Sensitivity measures were not significantly correlated (r = -.22, p = .11), indicating that the approach proposed in this study can indeed isolate independent sources of variance.

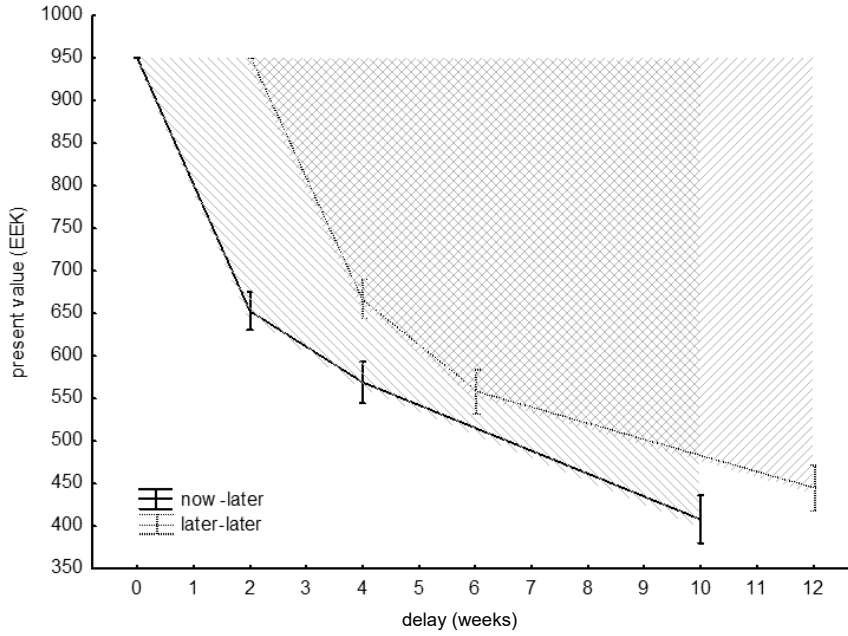


Figure 10.1 Discounting curves in the now-later and later-later conditions. The points depict subjective present values of 950 EEK delayed by 2, 4, and 10 weeks. Spreads denote standard errors. In the now-later condition, the present values were inferred from choices between delayed and immediate rewards. In the later-later condition, the values were inferred from choices between pairs of delayed rewards. The striped planes represent the areas over the curve (AOC) used to derive the Immediacy and Delay Sensitivity measures.

The variance captured by Delay Sensitivity was shared by the *k* and β parameters from the hyperboloid and quasi-hyperbolic models, respectively, which were virtually interchangeable (see Table 10.1). This cluster can be considered a representation of the traditional discounting rate. By contrast, Immediacy Sensitivity was largely independent of other discounting estimates, correlating only moderately with *s* and δ and being unrelated to the *k*, β, and the Delay Sensitivity cluster. Note that these findings are inconsistent with the beta-delta model, which assumes that the δ rather than the β component reflects the unitary discounting rate.

Table 10.1 Descriptive statistics and Spearman correlations of discounting parameters.

	M	*SD*	IS	*k*	*s*	δ	β
Delay Sensitivity (DS)	.00	1.00	-.22	.75***	.07	-.67***	-.68***
Immediacy Sensitivity (IS)	.00	1.00		.05	.27*	-.27*	-.01
k (hyperboloid)	.36	0.51			-.41***	-.30*	-.98***
s (hyperboloid)	.80	0.35				-.69***	.52***
δ (quasi-hyperbolic)	.93	0.05					.17
β (quasi-hyperbolic)	.78	0.18					

Note. $N = 56$. DS = Delay Sensitivity; IS = Immediacy Sensitivity; *k* = discounting parameter from Equation 1; *s* = time perception parameter from Equation 1; δ = discounting parameter from Equation 2; β = immediacy bias parameter from Equation 2.

$^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$

We analyzed the covariances among the personality and alcohol consumption measures (see Table 10.2) before using these as criterion variables for assessing the validity of discounting measures. In line with previous studies, the Sensation Seeking aspect was unrelated to the remaining and inter-related aspects of impulsivity (Duckworth & Kern, 2011). Also, as expected, participants who consumed most of their alcohol while binge-drinking ended up drinking more. Overall consumption also increased in tandem with problem frequency (no problems vs. some problems consumption difference, Cohen's $d = .71$; some problems vs. many problems, Cohens' $d = 1.23$; Median test $p < .001$). Meanwhile, the proportion of alcohol consumed while binging did not differ across problem levels (Median test $p = .27$).

Table 10.2 Descriptive statistics and Spearman correlations of self-reported and continuous alcohol consumption variables.

	M	*SD*	L_Plan	L_Pers	SS	Quantity	Binge
Urgency	52.4	11.1	.42**	.35**	.23	.31*	-.10
Lack of Planning	51.7	11.1		.49***	.22	.18	.14
Lack of Perseverance	54.6	10.2			-.07	.27	.01
Sensation Seeking	52.7	9.5				.34*	.24
Alcohol quantity	5.0	7.4					.53***
Binge drinking proportion	0.6	0.3					

Note. $N = 54$ for personality variables, $N = 47$ for alcohol variables. L_Plan = Lack of Planning; L_Pers = Lack of Perseverance; SS = Sensation seeking; Quantity = Alcohol quantity; Binge = Binge drinking proportion.

***$p < .001$, **$p < .01$, *$p < .05$

Table 10.3 Spearman correlations between discounting parameters and self-reported impulsivity and continuous alcohol consumption measures.

	DS	IS	*k*	*s*	δ	β
Urgency	.11	.28*	.19	.09	-.17	-.16
Lack of Planning	.08	.21	.02	.09	-.18	-.00
Lack of Perseverance	-.08	.34*	.01	.07	-.02	.02
Sensation Seeking	.07	.23	.18	-.01	-.17	-.16
Alcohol quantity	.27	.30*	.47***	-.16	-.21	-.44**
Binge drinking proportion	.28	.12	.31*	-.33*	.03	-.29*

Note. $N = 54$ for personality variables, $N = 47$ for alcohol variables. DS = Delay Sensitivity; IS = Immediacy Sensitivity; k = discounting parameter from Equation 1; s = time perception parameter from Equation 1; δ = discounting parameter from Equation 2; β = immediacy bias parameter from Equation 2.

***$p < .001$, **$p < .01$, *$p < .05$

Table 10.3 addresses the central question of this study—how different discounting measures relate to criterion variables. It first reveals that self-reported impulsivity was consistently related only to Immediacy Sensitivity (see also Figure 10.2). Given the high degree of overlap between different self-reported impulsivity measures, we also conducted a backward step-wise regression analysis to investigate which of the observed correlations were incremental (the "step" algorithm in R within least-squares regression using all four facets as predictors; all the involved measures were normally distributed). The best possible fit (adjusted $R^2 = .13$) was provided by a model including both Lack of Perseverance and Sensation Seeking (both $\beta = .03$, $p < .05$), indicating that these facets make at least partially independent contributions to Immediacy Sensitivity.

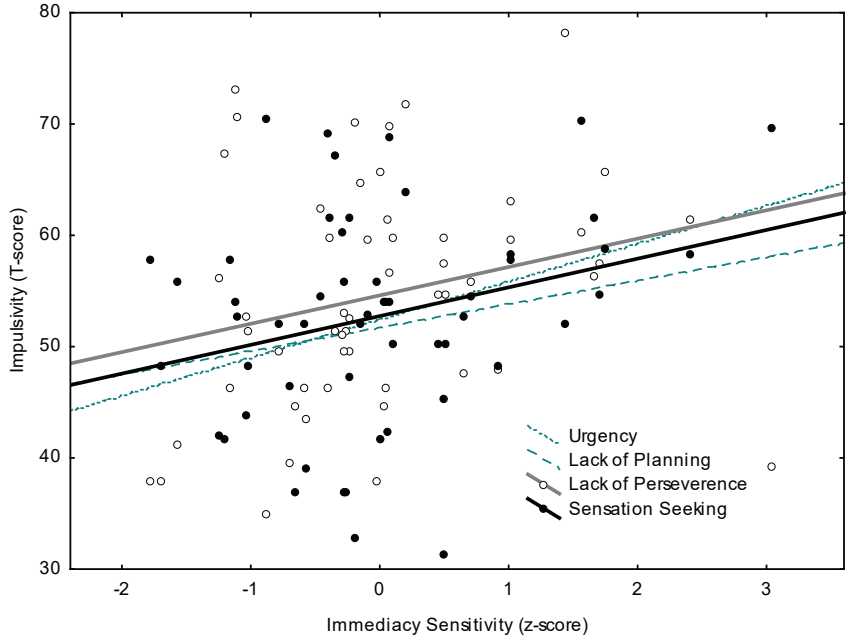


Figure 10.2 Correlations and regression lines associating Immediacy Sensitivity with different measures of self-reported Impulsivity. Data-points and solid lines represent significant correlations between Immediacy Sensitivity and Lack of Perseverance (filled circles) and Sensation Seeking (empty circles), respectively.

Unlike self-reported impulsivity, the measures of alcohol consumption were related to several discounting parameters. Participants with higher Immediacy Sensitivity as well as *k* and β tended to drink more. Binge drinking was related to *k* and β as well as *s*. Given that the discounting parameters were correlated, we again tested their incremental contributions. Due to violations of normality, robust regression analysis with M-estimation was used to predict alcohol quantity and binge proportion from Immediacy Sensitivity, *k*, and *s* (β was left out, due to its near-perfect correlation with *k*, see Table 10.1). All predictors were entered simultaneously. In these analyses, overall alcohol consumption was mainly a function of Immediacy sensitivity (β = 1.63, $p < .001$), as the effects of both the *s* and *k* did not reach significance ($p > .17$) in the regression analysis, despite demonstrating pair-wise relationships in the correlational analysis. Mean levels of Immediacy Sensitivity also differed near-significantly between drinking problem groups (no problems vs. some problems, Cohen's $d = .60$; some problems vs. many problems, Cohen's $d = .37$; Median test $p = .07$; other comparisons $p > .24$; see Figure 10.3). Meanwhile, binge drinking was significantly related only to the *s* parameter (β = -.28, $p < 05$; other effects $p > .15$).

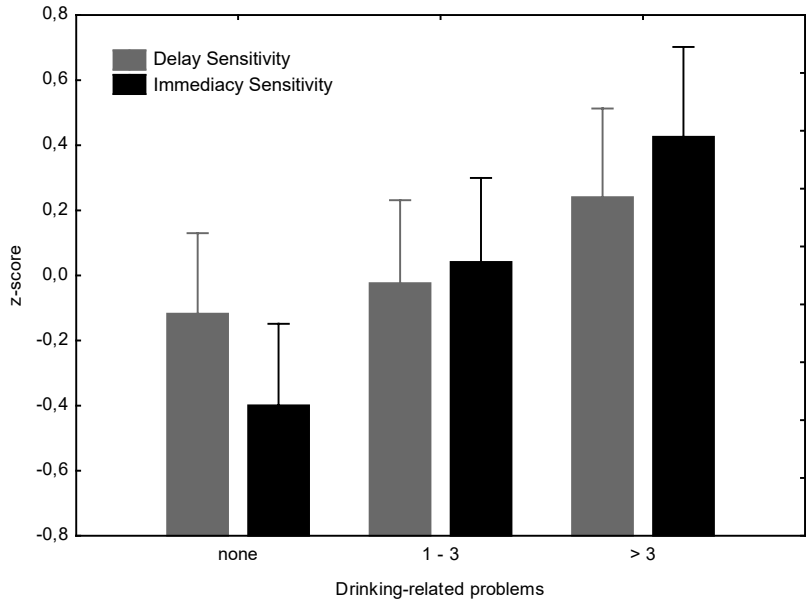


Figure 10.3 Comparison of Delay and Immediacy Sensitivity in predicting drinking-related problems. Only the differences in IS across groups with different levels of drinking-related problems were marginally significant at $p < .07$.

Discussion

Inspired by the neural correlates of inter-temporal decision-making (Berns et al., 2007; Metcalfe & Mischel, 1999; Peters & Büchel, 2011; Scheres et al., 2013), we distinguished between the Delay and Immediacy Sensitivity aspects of impulsivity. These aspects were operationalized by contrasting now-later and later-later choices in a delay discounting experiment. Our analyses revealed that Immediacy Sensitivity may outperform model-based alternatives, as well as Delay Sensitivity, in predicting self-report as well as behavioral impulsivity. As we explain in this section, these findings implicate the proposed operationalization of Immediacy Sensitivity as a promising step towards isolating the variance that is shared between delay discounting and impulsivity.

The consistent set of correlations involving Immediacy Sensitivity was, first of all, linked to the Lack of Perseverance and Sensation Seeking facets of self-reported impulsivity (Whiteside & Lynam, 2001). This pattern supports our hypothesis that Immediacy Sensitivity reflects processes related to the immediacy bias. The neural bases of Sensation Seeking can be traced back to the dopaminergic approach motivation system (DeYoung, 2013), which overlaps with brain areas involved in generating the motivational salience of immediate outcomes (Luo et al., 2009). Lack of Perseverance, meanwhile, is associated with the prefrontal substrates of self-control

(DeYoung et al., 2010), which are implicated in overcoming the immediacy bias (Essex et al., 2012). By contrast, the facets of impulsivity that did not correlate independently with Immediacy Sensitivity are associated with processes that are less relevant for generating and controlling the immediacy bias. Urgency is associated with avoidance motivation, which plays a lesser role in rewarding contexts (Whiteside & Lynam, 2001). Meanwhile, Lack of Planning captures inadequate consideration of the future, which should correlate with Delay rather than Immediacy Sensitivity.

Immediacy Sensitivity was also the best predictor of overall alcohol consumption and drinking-related problems. Even though the unitary discounting rate expressed in k or β also correlated with drinking quantity, these relationships did not remain significant after Immediacy Sensitivity was taken into account. In addition, Immediacy Sensitivity was the only discounting parameter to approach significance in predicting drinking-related problems. These findings imply that drinking variance has more to do with overcoming immediate temptations than considering the long-term adverse effects of alcohol consumption. This coincides with drinking motivation research indicating that alcohol consumption depends predominantly on the expected immediate rewards (such as excitement and belonging) rather than the postponed negative effects of alcohol intoxication, in particular among young consumers (Kuntsche, Knibbe, Gmel, & Engels, 2005).

Although the study was underpowered and the correlations involving Immediacy Sensitivity were mostly of medium size, their conceptual consistency suggests that the variance captured by this measure is meaningfully related to impulsivity. Our analyses also revealed that the Immediacy Sensitivity score outperformed more traditional model-based alternative methods for extracting trait-relevant parameters from discounting behavior. The pair of two-parameter models we analyzed (Laibson, 1997; Rachlin, 2006) both delivered a parameter that correlated with Delay Sensitivity and captured the traditional discounting rate. However, they failed to produce parameters that correlated with Immediacy Sensitivity or match its consistency in predicting criterion variables. This pattern is especially problematic for the beta-delta model, which aspires to disentangle the same processes as the present approach (Laibson, 1997). Specifically, the prediction that the β parameter reflects Immediacy Sensitivity while the δ parameter reflects Delay Sensitivity was not supported. Possibly, the model-free approach proposed here avoided losing variance by imposing the same model on all participants (Franck et al., 2015). In any case, our findings imply that the approach proposed here may be a more reliable method for realizing the ambition it shares with the beta-delta model.

As a small-scale test of a novel idea, this study has several limitations for future studies to overcome. Future studies should use larger and less homogenous samples. They could involve a wider selection of measures of self-report as well behavioral impulsivity. The experimental procedure could also include different (König, 2009) and real, instead of imaginary, rewards (Madden, Begotka, Raiff, & Kastern, 2003), as

well as delays (Lane, Cherek, Pietras, & Tcheremissine, 2003). Future studies could also examine the extent to which the enhanced validity of Immediacy Sensitivity stemmed from the mere doubling of data-points that were used to calculate it (8 compared to 4 for Delay Sensitivity, *k*, and *s,* and 3 for δ and β). They could also further investigate why the average discounting rates were not steeper in the now-later, compared to later-later, condition. Future research is also needed to map the relationships between Delay and Immediacy Sensitivity and other individual differences that are relevant for delay discounting (Green & Myerson, 2013; Peters & Büchel, 2011), such as general intelligence and working memory capacity (Shamosh & Gray, 2008), as well as prospection and simulation of future outcomes (Peters & Büchel, 2011). Finally, the finding that the timing-related *s* parameter from the hyperboloid model (Zauberman et al., 2009) predicted binge drinking could be followed up in relation to theories linking individual time perception differences to discounting as well as impulsivity (Wittmann & Paulus, 2008).

Conclusion

More broadly, the present results invite further scrutiny of the possibility that the traditional discounting rate represents a multitude of individual differences, only some of which are central to impulsivity. We related individual differences in valuation of delayed rewards to Delay Sensitivity and distinguished it from Immediacy Sensitivity, which we related to the control of the immediacy bias. Scores reflecting these aspects were derived using a model-free contrasting of now-later and later-later conditions of a delay discounting experiment. We found Immediacy Sensitivity, but not Delay Sensitivity or parameters of traditional models, to be a consistent correlate of self-reported impulsivity measures as well as of alcohol consumption. These findings suggest that the variance shared between discounting and impulsivity may involve an enhanced immediacy bias and/or a reduced ability to control this, rather than systematic differences in computing the present values of delayed rewards.

References

Allik, J. (2002). Towards a theory of personality. In L. Backman & C. von Hofsten (Eds.), *Psychology at the turn of the millennium, Vol 2: Social, developmental and clinical perspectives* (pp. 179–200). Hove, England: Psychology Press.

Ballard, K., & Knutson, B. (2009). Dissociable neural representations of future reward magnitude and delay during temporal discounting. *NeuroImage*, *45*, 143–150.

Bartra, O., McGuire, J. T., & Kable, J. W. (2013). The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage*, *76*, 412–427.

Beck, R. C., & Triplett, M. F. (2009). Test–retest reliability of a group-administered paper–pencil measure of delay discounting. *Experimental and Clinical Psychopharmacology*, *17*, 345–355.

Benhabib, J., Bisin, A., & Schotter, A. (2010). Present-bias, quasi-hyperbolic discounting, and fixed costs. *Games and Economic Behavior*, *69*, 205–223.

Berns, G. S., Laibson, D., & Loewenstein, G. (2007). Intertemporal choice: Toward an integrative framework. *Trends in Cognitive Sciences*, *11*, 482–488.

Bobova, L., Finn, P. R., Rickert, M. E., & Lucas, J. (2009). Disinhibitory psychopathology and delay discounting in alcohol dependence: Personality and cognitive correlates. *Experimental and Clinical Psychopharmacology*, *17*, 51–61.

Dawson, D. A., & Room, R. (2000). Towards agreement on ways to measure and report drinking patterns and alcohol-related problems in adult general population surveys: The Skarpö Conference overview. *Journal of Substance Abuse*, *12*, 1–21.

Dennis, J. E., Jr., Gay, D. M., & Walsh, R. E. (1981). An adaptive nonlinear least-squares algorithm. *ACM Transactions on Mathematical Software*, *7*, 348–368.

DeYoung, C. G. (2010). Toward a theory of the Big Five. *Psychological Inquiry: An International Journal for the Advancement of Psychological Theory*, *21*, 26–33.

DeYoung, C. G. (2013). The neuromodulator of exploration: A unifying theory of the role of dopamine in personality. *Frontiers in Human Neuroscience*, *7*, 762. doi:10.3389/fnhum.2013.00762

DeYoung, C. G., Hirsh, J. B., Shane, M. S., Papademetris, X., Rajeevan, N., & Gray, J. R. (2010). Testing predictions from personality neuroscience: Brain structure and the Big Five. *Psychological Science*, *21*, 820–828.

Doyle, J. R. (2013). Survey of time preference, delay discounting models. *Judgment and Decision Making*, *8*, 116–135.

Duckworth, A. L., & Kern, M. L. (2011). A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality*, *45*, 259–268.

Dzhafarov, E. N., Sekuler, R., & Allik, J. (1993). Detection of changes in speed and direction of motion: Reaction time analysis. *Perception & Psychophysics*, *54*, 733–750.

Essex, B. G., Clinton, S. A., Wonderley, L. R., & Zald, D. H. (2012). The impact of the posterior parietal and dorsolateral prefrontal cortices on the optimization of long-term versus immediate value. *The Journal of Neuroscience*, *32*, 15403–15413.

Figner, B., Knoch, D., Johnson, E. J., Krosch, A. R., Lisanby, S. H., Fehr, E., & Weber, E. U. (2010). Lateral prefrontal cortex and self-control in intertemporal choice. *Nature Neuroscience*, *13*, 538–539.

Franck, C. T., Koffarnus, M. N., House, L. L., & Bickel, W. K. (2015). Accurate characterization of delay discounting: A multiple model approach using approximate Bayesian model selection and a unified discounting measure. *Journal of the Experimental Analysis of Behavior*, *103*, 218–233.

Green, L., & Myerson, J. (2004). A discounting framework for choice with delayed and probabilistic rewards. *Psychological Bulletin*, *130*, 769–792.

Green, L., & Myerson, J. (2013). How many impulsivities? A discounting perspective. *Journal of the Experimental Analysis of Behavior*, *99*, 3–13.

Green, L., Myerson, J., & Macaux, E. W. (2005). Temporal discounting when the choice is between two delayed rewards. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1121–1133.

Hirsh, J. B., Morisano, D., & Peterson, J. B. (2008). Delay discounting: Interactions between personality and cognitive ability. *Journal of Research in Personality*, *42*, 1646–1650.

Johnson, M. W., & Bickel, W. K. (2008). An algorithm for identifying nonsystematic delay-discounting data. *Experimental and Clinical Psychopharmacology*, *16*, 264–274.

Kable, J. W., & Glimcher, P. W. (2007). The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience*, *10*, 1625–1633.

Kirby, K. N. (2009). One-year temporal stability of delay-discount rates. *Psychonomic Bulletin & Review*, *16*, 457–462.

Kirby, K. N., & Herrnstein, R. J. (1995). Preference reversals due to myopic discounting of delayed reward. *Psychological Science*, *6*, 83–89.

König, C. J. (2009). A generalizability study of time discounting: Some people react more strongly to domain differences than others. *Swiss Journal of Psychology*, *68*, 43–50.

Kuntsche, E., Knibbe, R., Gmel, G., & Engels, R. (2005). Why do young people drink? A review of drinking motives. *Clinical Psychology Review*, *25*, 841–861.

Kuntsche, E., Knibbe, R., Gmel, G., & Engels, R. (2006). Who drinks and why? A review of socio-demographic, personality, and contextual issues behind the drinking motives in young people. *Addictive Behaviors*, *31*, 1844–1857.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, *112*, 443–477.

Lane, S. D., Cherek, D. R., Pietras, C. J., & Tcheremissine, O. V. (2003). Measurement of delay discounting using trial-by-trial consequences. *Behavioural Processes*, *64*, 287–303.

Levy, D. J., & Glimcher, P. W. (2012). The root of all value: A neural common currency for choice. *Current Opinion in Neurobiology*, *22*, 1027–1038.

Liu, L., Feng, T., Wang, J., & Li, H. (2012). The neural dissociation of subjective valuation from choice processes in intertemporal choice. *Behavioural Brain Research*, *231*, 40–47.

Luhmann, C. C. (2009). Temporal decision-making: insights from cognitive neuroscience. *Frontiers in Behavioral Neuroscience*, *3*, 39. doi:10.3389/neuro.08.039.2009

Luhmann, C. C. (2013). Discounting of delayed rewards is not hyperbolic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1274–1279.

Luo, S., Ainslie, G., Giragosian, L., & Monterosso, J. R. (2009). Behavioral and neural evidence of incentive bias for immediate rewards relative to preference-matched delayed rewards. *The Journal of Neuroscience*, *29*, 14820–14827.

MacKillop, J., Amlung, M. T., Few, L. R., Ray, L. A., Sweet, L. H., & Munafò, M. R. (2011). Delayed reward discounting and addictive behavior: A meta-analysis. *Psychopharmacology*, *216*, 305–321.

Madden, G. J., Begotka, A. M., Raiff, B. R., & Kastern, L. L. (2003). Delay discounting of real and hypothetical rewards. *Experimental and Clinical Psychopharmacology*, *11*, 139–145.

Mahalingam, V., Stillwell, D., Kosinski, M., Rust, J., & Kogan, A. (2014). Who can wait for the future? A personality perspective. *Social Psychological and Personality Science*, *5*, 573–583.

Manning, J., Hedden, T., Wickens, N., Whitfield-Gabrieli, S., Prelec, D., & Gabrieli, J. D. E. (2014). Personality influences temporal discounting preferences: Behavioral and brain evidence. *NeuroImage*, *98*, 42–49.

McClure, S. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2004). Separate neural systems value immediate and delayed monetary rewards. *Science*, *306*, 503–507.

McKerchar, T. L., Green, L., Myerson, J., Pickford, T. S., Hill, J. C., & Stout, S. C. (2009). A comparison of four models of delay discounting in humans. *Behavioural Processes*, *81*, 256–259.

Metcalfe, J., & Mischel, W. (1999). A hot/cool-system analysis of delay of gratification: Dynamics of willpower. *Psychological Review*, *106*, 3–19.

Mõttus, R., Pullmann, H., & Allik, J. (2006). Toward more readable big five personality inventories. *European Journal of Psychological Assessment*, *22*, 149–157.
Myerson, J., & Green, L. (1995). Discounting of delayed rewards: Models of individual choice. *Journal of the Experimental Analysis of Behavior*, *64*, 263–276.
Myerson, J., Green, L., & Warusawitharana, M. (2001). Area under the curve as a measure of discounting. *Journal of the Experimental Analysis of Behavior*, *76*, 235–243.
Odum, A. L. (2011). Delay discounting: Trait variable? *Behavioural Processes*, *87*, 1–9.
Ostaszewski, P. (1996). The relation between temperament and rate of temporal discounting. *European Journal of Personality*, *10*, 161–172.
Peters, J., & Büchel, C. (2011). The neural mechanisms of inter-temporal decision-making: Understanding variability. *Trends in Cognitive Sciences*, *15*, 227–239.
Phelps, E. S., & Pollak, R. A. (1968). On second-best national saving and game-equilibrium growth. *The Review of Economic Studies*, *35*, 185–199.
R Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/
Rachlin, H. (2006). Notes on discounting. *Journal of the Experimental Analysis of Behavior*, *85*, 425–435.
Reynolds, B. (2006). A review of delay-discounting research with humans: Relations to drug use and gambling. *Behavioural Pharmacology*, *17*, 651–667.
Reynolds, B., Ortengren, A., Richards, J. B., & de Wit, H. (2006). Dimensions of impulsive behavior: Personality and behavioral measures. *Personality and Individual Differences*, *40*, 305–315.
Rossow, I. (2008). Alcohol consumption and discounting. *Addiction Research & Theory*, *16*, 572–584.
Scheres, A., de Water, E., & Mies, G. W. (2013). The neural correlates of temporal reward discounting. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*, 523–545.
Sellitto, M., Ciaramelli, E., & di Pellegrino, G. (2011). The neurobiology of intertemporal choice: Insight from imaging and lesion studies. *Reviews in the Neurosciences*, *22*, 565–574.
Shamosh, N., & Gray, J. (2008). Delay discounting and intelligence: A meta-analysis. *Intelligence*, *36*, 289–305.
Sharma, L., Markon, K. E., & Clark, L. A. (2014). Toward a theory of distinct types of "impulsive" behaviors: A meta-analysis of self-report and behavioral measures. *Psychological Bulletin*, *140*, 374–408.
Smith, C. L., & Hantula, D. A. (2008). Methodological considerations in the study of delay discounting in intertemporal choice: A comparison of tasks and modes. *Behavior Research Methods*, *40*, 940–953.
Sripada, C. S., Gonzalez, R., Luan Phan, K., & Liberzon, I. (2011). The neural correlates of intertemporal decision-making: Contributions of subjective value, stimulus type, and trait impulsivity. *Human Brain Mapping*, *32*, 1637–1648.
Stautz, K., & Cooper, A. (2013). Impulsivity-related personality traits and adolescent alcohol use: A meta-analytic review. *Clinical Psychology Review*, *33*, 574–592.
Story, G., Vlaev, I., Seymour, B., Darzi, A., & Dolan, R. (2014). Does temporal discounting explain unhealthy behavior? A systematic review and reinforcement learning perspective. *Frontiers in Behavioral Neuroscience*, *8*, 76. doi:10.3389/fnbeh.2014.00076
Swann, A. C., Bjork, J. M., Moeller, F. G., & Dougherty, D. M. (2002). Two models of impulsivity: Relationship to personality traits and psychopathology. *Biological Psychiatry*, *51*, 988–994.

Takahashi, T. (2009). Theoretical frameworks for neuroeconomics of intertemporal choice. *Journal of Neuroscience, Psychology, and Economics*, *2*, 75–90.

Whiteside, S. P., & Lynam, D. R. (2001). The Five Factor Model and impulsivity: using A structural model of personality to understand impulsivity. *Personality and Individual Differences*, *30*, 669–689.

Wittmann, M., Leland, D., & Paulus, M. (2007). Time and decision making: Differential contribution of the posterior insular cortex and the striatum during a delay discounting task. *Experimental Brain Research*, *179*, 643–653.

Wittmann, M., & Paulus, M. P. (2008). Decision making, impulsivity and time perception. *Trends in Cognitive Sciences*, *12*, 7–12.

Zauberman, G., Kim, B. K., Malkoc, S. A., & Bettman, J. R. (2009). Discounting time and time discounting: Subjective time perception and intertemporal preferences. *Journal of Marketing Research*, *46*, 543–556.

11 TOWARDS MEANINGFUL COMPARISONS OF PERSONALITY IN LARGE-SCALE CROSS-CULTURAL STUDIES

GWEN GARDINER, KYLE SAUERBERGER,
MEMBERS OF THE INTERNATIONAL SITUATIONS PROJECT,
AND DAVID FUNDER

One of Jüri Allik's major, pioneering contributions to psychology is the assessment of personality across numerous cultures. His contributions have inspired many other large collaborations of international researchers to move beyond early work confirming the Five Factor Model cross-culturally to assessing the reliability and validity of a broad range of personality traits. Cross-cultural comparisons of personality traits may be problematic if measures have unique meanings in different cultural contexts that influence how individuals respond to items. In this chapter we present a new and relatively simple method for assessing the comparability of measures in large-scale cross-cultural studies, and illustrate the method using responses to the Big Five Inventory-2 (BFI-2) from 15,368 participants in 63 countries participating in the International Situations Project.

Introduction

The recent growth in cross-cultural research has brought with it an expansion of the study of personality across cultures, particularly with large collaborations of researchers accumulating data across numerous cultural groups. Jüri Allik has been a pioneering participant and leader in this effort, and his contributions to the understanding of personality across cultures is one—just one—of his significant career accomplishments (e.g., Allik & McCrae, 2004; Mõttus, Allik, & Realo, 2010; Schmitt et al., 2007).

Initially, most cross-cultural research on personality focused on testing if the Five Factor Model (FFM) of personality was reproducible in samples outside of the

The present research was supported by the US National Science Foundation under Grant BCS-1528131. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the individual researchers and do not necessarily reflect the views of the National Science Foundation. Data gathering in the Czech Republic was supported by grant 17-14387S by the Czech Science Foundation and by institutional research funding RVO: 68081740 from the Institute of Psychology, Czech Academy of Sciences.
Members of the International Situations Project contributed the data used in the analyses reported in this chapter. Their names and affiliations are listed in Appendix 11A.

E-mail: ggard001@ucr.edu, kylesauerberger@gmail.com, and funder@ucr.edu

Western world (with the answer being, "generally, yes"). Later, researchers expanded the research question to include the reliability or accuracy of personality profiles of cultures. This has led to the issue of whether measures can be compared across cultural groups, who may have unique interpretations of the items in the measures. Unique cultural interpretations of items could bias results and limit the conclusions that can be drawn from the data (see, e.g., Allik & Realo, 2017).

Various methodological approaches have been suggested and used in an attempt to detect and (perhaps) correct for cultural biases in responses to measurement instruments and the study of "measurement invariance" has become a complex and daunting statistical issue (e.g., van de Schoot, Lugtig, & Hox, 2012). In the present chapter we suggest and demonstrate a new and relatively simple approach to assessing the comparability of measures in large-scale cross-cultural studies.

The Problem of Cross-Cultural Comparability

Early work on assessing personality around the world typically tested the generalizability of the FFM in one or two non-Western societies (e.g., Gurven, von Rueden, Massenkoff, Kaplan, & Lero Vie, 2013) or compared personality trait relationships and behavioral expressions among a handful of diverse nations (e.g., Ching et al., 2014). While each individual study provides unique contributions, the most informative studies are those that assess a wide range of cultures (Allik & Realo, 2017). A large sample of cultures is more informative in the same way a large sample of individuals is more informative. The large sample of cultures will exhibit a wider range of traits and be more representative of the larger population. Additionally, researchers interested in the reliability of country trait profiles need large, independent samples with enough overlapping cultures to test the replicability of previous findings (Allik & Realo, 2017). The number of large-scale cross-cultural research projects will continue to grow in the coming years as more researchers form international collaborations and technological access expands around the world allowing for easier data collection in more diverse nations.

One crucial aspect to cross-cultural research is assessing the comparability of the measures used across a range of diverse cultural groups. Typical questionnaires used to measure how a specific construct varies across cultures may inadvertently assess other cultural characteristics in relation to responding to the questionnaire itself. For example, a tendency to always choose the most extreme responses on a Likert scale biases the overall score on that measure. Response styles to questionnaires have been linked to cultural dimensions, implying that any cultural differences found in questionnaire results are partially due to cultural differences in responding to surveys (Harzing, 2006). Additionally, in cross-cultural studies researchers typically translate existing measures into the native language of the assessment group. Items that are

mistranslated or represent a distinct cultural construct that is not universal will also bias the overall results from measures (Chen, 2008).

Given the range of potential sources of bias in the data, researchers have developed methods for testing the comparability of measures across groups. Typically, the factor model of a measure is compared between a reference group and a comparison group (Byrne & van de Vijver, 2010). The comparability of a measure is determined by the fit statistics of the model, often using seemingly arbitrary thresholds for determining "good fit." This method is problematic for researchers interested in understanding the nuances in potential cultural biases in the data, because it provides only a single overall measure of fit, without indicating clearly which items on measures are the source of convergence and difference without further testing. Additionally, the traditional method of comparing each new cultural group to a reference group, often the United States (US), becomes exponentially difficult as the number of countries grows and more comparisons are needed (Byrne & van de Vijver, 2010).

Large-scale Cross-Cultural Assessments of Personality

While the number of large-scale cross-cultural assessments of personality is growing but still small, the range of methods used to test if meaningful comparisons can be made across cultures is wide. Formal statistical models, although available, are also difficult to understand and use, and their application to actual cross-cultural data remains rare (although see Zecca et al., 2013 for an exception). Instead, the most common method is to compare the country level trait scores with previously collected country trait scores, and also with other country level data. Convergence across samples and associations with independently-measured country-level measures (such as demographic or economic development information) implies that the variation in personality scores across cultures is meaningful (Mõttus et al., 2010). While some external country level predictors of aggregated personality traits are surprising (e.g., Heine, Buchtel, & Norenzayan, 2008), it is probably still too soon to determine the validity of this method (Mõttus et al., 2010).

McCrae, Terracciano and colleagues (2005) were among of the first researchers to collect data on personality traits across a wide range of countries that had been previously assessed, allowing the replicability of results to be examined. Previous cross-cultural comparisons of personality traits had involved secondary data analysis accumulated from multiple independent research projects which, while maximizing the number of countries that could be compared, limited the degree to which the findings across cultures could be considered directly comparable. In an important advance over that approach, McCrae and colleagues (2005) assessed personality traits using the NEO PI-R in 50 cultures by asking college students to rate the personality of someone they knew well. Observer reports were used to limit biases inherent in self-

reports and potentially expand the representativeness of the sample beyond traditional college students. Because McCrae and colleagues (2005) was one of the first large-scale assessments of personality, using the same measure in different cultures, the researchers were also one of the first to attempt to assess the comparability of their measure across numerous cultural groups. First, the researchers pooled all the data together and tested the Big Five factor structure using confirmatory factor analysis (CFA). Then, they used Procrustes rotation to compare the factor structure of each culture with the US as a reference group and found evidence for comparability across the groups, with some exceptions in the African countries.

Schmitt, Allik, and colleagues (2007) followed a similar method of testing the comparability of their country level traits score, this time assessed using the 44-item Big Five Inventory (BFI; John & Srivastava, 1999). The factor structure of the BFI was first examined in the total sample of the study and the authors found good fit to the data. To test for cultural differences in the factor structure, the countries were grouped into 10 regions that were then compared with the US as a reference group using Procrustes rotation. Overall, the researchers found evidence for good congruence. The large number of countries overlapping between Schmitt and colleagues (2007) and McCrae and colleagues (2005) allowed for the reliability of country level trait scores to be assessed using different measures. The correlations of personality traits between samples was positive for all traits but only statistically significant at $p < .05$ for extraversion, conscientiousness, and neuroticism. The moderate evidence found for the reliability of the country trait scores strengthened the argument that the cross-cultural variation in personality measures assesses something meaningful, rather than random noise.

Along the same lines as Schmitt et al. (2007), Bartram (2013) assessed the accuracy of personality trait measures of countries by correlating them with findings from previous studies and with other country-level variables. Once again, a different measure of personality was used, providing more evidence for convergent validity of the trait averages. The Occupational Personality Questionnaire (OPQ32) is a personality assessment questionnaire used for studies in the workforce and was tested in 31 countries. Items from the measure were selected to represent the Big Five traits. The OPQ32 is a forced choice assessment in which participants must choose from a list of 4 characteristics an item that is most like them and an item that is least like them. Forced-choice measures are especially useful in cross-cultural comparisons because they can decrease the effects of response styles, a tendency to bias results that are linked to some cultural aspects (Harzing, 2006). However, forced-choice measures can become problematic for traditional statistical tests of equivalence that assume item independence, which may be one reason no formal tests of equivalence were reported (Bartram, 2013).

Thus far, cross-cultural assessments of personality using a large number of cultural groups have largely focused on confirming the factor structure within each group

and the convergence of scores with previous, independent assessments of the same construct. Thalmayer and Saucier (2014) assessed the QB6, a measure of the Big Six that can be reduced to the Big Five, across 26 countries. The countries were separated into three groups which were used to independently verify the factor model. Using "domain specific" fit statistics thresholds for multivariate measures derived from Hopwood and Donnellan (2010), the researchers found good model fit for the factor structure and item loadings for both the Big Five and the Big Six. However, even with the lower domain specific thresholds, removing problematic items, and excluding countries, the researchers still did not have enough evidence for equality in variable intercepts, a step usually considered necessarily for comparing means across groups. The researchers subsequently cautioned against group mean comparisons and did not report any trait scores for the countries assessed (Thalmayer & Saucier, 2014).

In sum, current methods for testing the accuracy of personality trait scores at the country level have been quite limited. The most common method is to compare newly assessed country trait scores with previously collected country trait scores to determine the reliability and validity of the findings. The few attempts at more formal methods have found evidence for the comparability of the measures across groups when using simplified methods for testing the factor structure (e.g., McCrae et al., 2005; Schmitt et al., 2007) and limited evidence when tested with more traditional psychometric methods (e.g., Thalmayer & Saucier, 2014). Recommendations for modifying existing methods for a large number of groups are labor intensive and lack the ability to compare numerous cultures to each other, rather than solely to one reference group. Therefore, a new, simpler approach might be worth trying, one that does not incorporate strict or arbitrary statistical thresholds for success while still allowing researchers flexibility for discovering potentially problematic items or cultural groups in their data.

The Comparability of Measures Using an Inter-item Correlation Matrix

A critical concern for researchers interested in knowing whether or not a measure has comparable meaning across cultural groups is the degree to which the items on the measure are understood the same way. Only to the degree that items on the measure have the same meaning to the individuals who respond to them can we infer that different responses to the items reflect differences in the construct the researcher is trying to assess. The key idea underlying the method proposed in this chapter is simply this: The meaning of each item on a psychological measurement instrument can be conceptualized in terms of its relationships with the other items in the measure.

This approach to item meaning is analogous to how words are defined in a dictionary—each word is defined using other words in the dictionary, which in turn

are defined by using still other words in the dictionary. The underlying assumption is that the meaning of a word is fully contained in, and reflected by, its relation to other words. Analogously, the meaning of an item on a self-report scale, especially one with a large number of items, could be assumed to be reflected in its relationships to the other items in the scale. A complete item-by-item correlation matrix, then, could be taken to reflect the meaning of each item in terms of its relationships with all of the others, and the overall pattern of correlations to reflect the meaning of the measure as whole.[1]

Therefore, one possible method for assessing the degree to which participants from different countries infer similar meaning from the items on a scale is by calculating the relationships between each item and every other item within each country. Each country will have its own resulting matrix of inter-item correlations that can then be correlated with the inter-item correlation matrix of every other country. The resulting correlation between any two countries (which is simply the vector correlation between the two sets of non-redundant inter-item correlations) represents how similarly participants in the two countries interpret each item in relation to every other item. In a study of many countries, this approach can be expanded to produce a country-by-country matrix that reveals how similar the pattern of inter-item correlations is between any two countries in the sample, how similar the pattern within any given country is to the average pattern of other countries, and how similar the pattern of item meaning is overall, across the world.

Comparing inter-item correlation matrices has several possible benefits over traditional methods of testing for the comparability of measures across cultures. First, it is simple and transparent. Compare this method to the one illustrated by Davidov, Schmidt, and Schwartz (2008). Their sophisticated approach began with computing a confirmatory factor analysis (CFA) within each country in their sample (20 countries), attempting to derive a factor structure adequate to describe all of the countries' response patterns, then following up with a multigroup confirmatory factor analysis (MGCFA) to assess the degree to which this attempt was successful. We suggest that our method is a much simpler and more transparent way to assess configural invariance. A second advantage is that our method clearly shows the degree to which each country is similar (or dissimilar) in its configural structure to each other country, and also, the degree which it is similar and dissimilar to other countries overall—information which the conventional MGCFA does not so readily provide.

Thus, researchers can compare all countries with each other, rather than every country with a single reference country, such as (most often) the US. This capability allows researchers to see if countries that have a lower correlation with the US also have a lower correlation with many other countries, indicating random error in the data, or if they are more similar to other culturally comparable countries,

[1] Conventional factor analytic methods are rooted in this item-by-item matrix and derive all of their information from it, but focus on latent factors or other multi-variate constructs that emerge, rather than the matrix itself.

implying a cultural bias in the data. Inter-item correlation matrices also can work well for multivariate measures, such as Big Five personality measures, which can be problematic for traditional methods of comparison (Hopwood & Donnellan, 2010). The simple method of matrix comparison is also useful even if a measure does not have any strong latent variables or has excess items that do not correspond to specific constructs, because it is the meaning of each item that is assessed, rather than latent variables that may or may not be culturally relevant for all groups tested. Lastly, this method is easier to conduct than traditional methods that require expensive software or advanced statistical knowledge to perform and understand, which can and we suspect does often limit the use of these methods in the field.[2] Here we present an example of this new method using personality data collected as part of a large-scale international research project.

Method

Participants

The International Situations Project (ISP) is a large, international collaboration involving over 130 researchers representing 63 countries and 40 languages (see Table 11.1). Participants (N = 15,368) were recruited by collaborators at their local university to answer a survey online that included several measures of personality, values, and situational experience. All measures were first translated into the local language and then back-translated by an independent source. The back-translation and original English were compared, and any discrepancies resolved.

Measures

Personality was assessed using the Big Five Inventory-2 (BFI-2; Soto & John, 2017). The BFI-2 consists of 60 items that measure the Big Five traits—Extraversion, Agreeableness, Conscientiousness, Negative Emotionality, and Open-Mindedness—and 15 facets (three facets nested within each Big Five trait). In the present analyses, we shall focus on the 60 items rather than their subsuming traits or facets.

Results and Discussion

An inter-item correlation matrix was first created for each country by correlating every BFI-2 item with every other BFI-2 item, resulting in 60 x 60 item matrix for each of 63 countries.[3] Then, each country's inter-item correlation matrix was correlated with

[2] The analyses reported in this paper were conducted using the open-source program R (R Core Team, 2017) and required no specialized or proprietary software.

[3] The number of non-redundant correlations in this matrix is (60 x 59)/2, or 1,770, and these are the correlations that enter into the vector correlations that compare each pair of countries.

Table 11.1 Demographic information and sample size by country.

Country	Mean age	Total *N*	% female	Country	Mean age	Total *N*	% female
Argentina	24.83	140	78.85	Mexico	23.88	247	58.37
Australia	19.84	196	76.02	Netherlands	20.13	301	81.33
Austria	21.26	113	81.42	New Zealand	19.19	129	86.05
Belgium	19.14	50	84.00	Nigeria	24.75	135	33.58
Bolivia	21.01	135	57.78	Norway	23.89	159	74.21
Brazil	23.68	310	72.17	Pakistan	20.61	114	50.00
Bulgaria	25.05	152	70.67	Palestine	22.17	295	83.39
Canada	21.86	304	79.14	Peru	28.21	74	58.26
Chile	21.45	386	66.41	Philippines	19.71	337	69.18
China	25.31	432	46.01	Poland	22.35	234	83.33
Colombia	21.68	181	74.03	Portugal	21.66	157	87.82
Croatia	21.46	218	64.68	Romania	22.84	177	57.06
Czech Republic	22.65	193	80.83	Russia	21.92	159	78.48
Denmark	22.94	246	79.92	Senegal	23.32	635	47.48
Estonia	25.88	293	83.96	Serbia	23.57	185	75.85
France	22.60	231	85.53	Singapore	20.93	136	77.94
Georgia	20.29	140	80.00	Slovakia	22.41	148	69.59
Germany	24.49	458	75.70	Slovenia	20.43	123	57.38
Greece	24.09	225	79.22	South Africa	22.21	256	66.67
Hong Kong	19.00	144	59.15	South Korea	22.35	281	58.36
Hungary	25.33	178	66.67	Spain	19.73	419	85.20
India	24.99	221	57.04	Sweden	†	130	72.22
Indonesia	21.85	131	52.71	Switzerland	22.45	755	84.30
Israel	25.35	173	61.40	Taiwan	19.71	162	76.54
Italy	21.86	717	64.57	Thailand	19.24	196	80.32
Japan	22.58	243	61.98	Turkey	21.09	329	68.29
Jordan	19.87	141	80.85	Uganda	22.63	93	64.52
Kenya	21.17	139	65.47	Ukraine	23.91	244	75.79
Latvia	24.87	169	82.84	United Kingdom	25.61	136	88.41
Lithuania	20.26	145	78.47	United States	19.85	1366	67.72
Macedonia	21.22	54	74.07	Vietnam	19.05	168	77.25
Malaysia	21.53	230	71.05	*World sample*	*22.34*	*15,368*	*70.13*

Note: † = Data not available.

every other country's correlation matrix, resulting in a 63 x 63 correlation matrix (please see Table 11.S1)[4]. The resulting correlations between countries represent the degree of similarity in how items are responded to in terms of other items, with higher numbers indicating more similarity between countries. Additionally, the average correlation between each country's inter-item matrix and the matrices of the other countries was calculated, to determine which countries have the greatest overall similarity with other countries (see Table 11.2). Among the participants in the ISP, the country that is most similar to every other country is, unsurprisingly, the US[5] (r = .80), followed by Switzerland (r = .77), Canada (r = .77), Estonia (r = .77), and the Philippines (r = .77). The least similar countries, meaning countries in which participants interpreted items the most distinctively compared to the other countries, were Macedonia (r = .46), Pakistan (r = .50), Uganda (r = .51), Vietnam (r = .53), and Indonesia (r = .55). Overall, the average inter-item matrix correlation among countries was r = .69.

The overall matrix in Table 11.S1 also allows researchers to easily compare countries that are the most similar and the least similar. In these data, the countries that are most similar to each other are the US and Canada (r = .91), the US and the Philippines (r = .91), and Germany and Switzerland (r = .91). The countries that are least similar to each other are Uganda and Macedonia (r = .34), Vietnam and Macedonia (r = .35), and Belgium and Uganda (r = .35). The low comparability correlations for Macedonia and Belgium might reflect the smaller sample size for those countries.

Researchers can also test if low average correlations for countries are consistent across all countries or vary according to cultural differences. For example, it is possible to test if countries with lower overall correlations have equally low correlations with other culturally similar countries with lower overall countries. This shows whether low correlations are the result of random error in the data or if it reflects some underlying cultural bias. For example, one of the least similar countries overall is Pakistan (r = .50). However, the inter-item matrix correlation between Pakistan and India, a geographically and culturally close country, is one of the highest country correlations for Pakistan (r = .59). Uganda, another country with a low average correlation but less culturally similar to Pakistan than India, has a relatively low matrix correlation with Pakistan (r = .42). However, Uganda has a higher matrix correlation with Kenya (r = .62) and Nigeria (r = .60), other African countries in the dataset, and a lower matrix correlation with Vietnam (r = .45), a country culturally distinct from Uganda. Thus, while overall Pakistan, India, Uganda, Kenya, and Nigeria are all dissimilar to others, Pakistan and India are more similar in their dissimilarity compared with other countries and Uganda, Kenya, and Nigeria are also more similar to each other in their dissimilarities than with other countries.

---

[4] Table 11.S1 is too large to appear in print, but can be accessed via a Google Sheet at https://goo.gl/rNynoq

[5] The BFI-2 was originally developed in the US.

Table 11.2 Average similarity of inter-item matrix correlations of the BFI-2 items, by country.

Country	Average	Country	Average
United States	0.80	Slovakia	0.71
Canada	0.77	Sweden	0.71
Estonia	0.77	Argentina	0.70
Philippines	0.77	Austria	0.70
Switzerland	0.77	Denmark	0.70
Chile	0.76	France	0.70
Germany	0.76	Hong Kong	0.70
Turkey	0.76	Israel	0.70
Croatia	0.75	New Zealand	0.70
South Africa	0.75	Peru	0.70
Hungary	0.74	Portugal	0.70
Mexico	0.74	Russia	0.70
Serbia	0.74	Jordan	0.69
Spain	0.74	Lithuania	0.69
United Kingdom	0.74	Palestine	0.69
China	0.73	Slovenia	0.69
Italy	0.73	Latvia	0.68
Netherlands	0.73	Thailand	0.68
Romania	0.73	Bulgaria	0.67
Australia	0.72	Georgia	0.65
Brazil	0.72	India	0.65
Czech Republic	0.72	Nigeria	0.64
Japan	0.72	Kenya	0.62
Norway	0.72	Belgium	0.60
Poland	0.72	Malaysia	0.60
Singapore	0.72	Senegal	0.60
South Korea	0.72	Indonesia	0.55
Taiwan	0.72	Vietnam	0.53
Ukraine	0.72	Uganda	0.51
Bolivia	0.71	Pakistan	0.50
Colombia	0.71	Macedonia	0.46
Greece	0.71	*World average*	*0.69*

One potential problem associated with this method is the lack of a metric to judge the resulting correlations among countries. This difficulty is not unique to this method; for more complex methods in the literature, various thresholds for acceptable degrees of "measurement invariance" have been proposed without clear justification. In the present case, as well, it is not obvious what a "good" correlation between two countries' inter-item matrices is, that implies sufficient comparability of the measure across these countries. One way to generate a reference point is by comparing the actual results with randomized correlations among arbitrary groups. In other words, what if it truly did not matter, at all, what country a participant was from? To test this hypothetical possibility, we removed the country identification from each of our more than 15,000 participants, and then re-assigned them to pseudo-"countries," randomly.

Specifically, a randomization program assigned each of the more than 15,000 participants to one of 63 groups, weighted to have equal sample sizes with the countries in the original dataset. Then, new inter-item correlation matrices were calculated for each of the 63 randomized groups. These group correlation matrices were then correlated with every other group to form a new inter-item correlation similarity matrix among randomized groups. The resulting average correlation among all the randomized cultural groups was $r = .80$, which is higher than the average correlation among the actual countries ($r = .69$). The correlation coefficient generated from randomized groups represents the upper limit of the best inter-item correlation matrix that can be expected from the data, given no cultural biases in item responses.

Once again, however, it is difficult to determine a metric for what is considered a high enough or too low of a matrix correlation for researchers to conclude enough equality in item interpretation for measures to be reliably compared across cultural groups. One method for assessing the amount of discrepancy expected is to assess the similarity of inter-item matrices within subgroups of one culture assumed to have very little, if any, discrepancies among groups. Six different sites within the US collected data for the ISP, representing Alabama, California, Connecticut, Idaho, Illinois, and Texas. While personality traits vary across the states (see Rentfrow, Gosling, & Potter, 2008), it is generally assumed that the comparability of measures across states is not an issue. Therefore, the US states represent a baseline metric for expected discrepancies between randomized group inter-item matrices and actual group inter-item matrices.

Following the same method as before, an inter-item correlation matrix was calculated for each of the six US sites and then correlated with each of the other US sites. The average inter-item correlation matrix for US sites was $r = .83$. For the randomized US sites, each US participant was randomly assigned to one of six groups, weighted to match the sample size of the original US sites. An inter-item correlation matrix was calculated for each of the randomized groups and then correlated with each of the other randomized groups. The average inter-item correlation matrix for the randomized US sites was $r = .84$, implying US states do not impact the comparability of measures across groups any more than randomly assigned groups.

Conclusion

Jüri Allik's pioneering contributions to cross-cultural research opened up new possibilities and new methodological challenges for psychology. One of those challenges is how to separate real differences between cultures from those that are products of response bias, shifts in meaning, or other measurement artifacts. The present chapter presents a new approach to that challenge and the current demonstration of the approach provides some interesting and important new information but also, as always, leaves us wanting to learn more.

The new information is the possibly-encouraging finding that the average similarity of patterns of item response to the BFI-2 across 63 countries is $r = .69$. However, we say "possibly encouraging" because we lack a clear benchmark of comparison. An r of .69 is generally regarded as large in most research contexts[6], but our randomization procedure described in this chapter suggests that if country really did not matter for item response, the r would be .80. A further analysis found that patterns of item response were consistent across several states of the US, with the actual and pseudo-groupings resulting in almost exactly equivalent patterns of response similarity. So our overall conclusion is this: For responding to the items of the BFI-2, it does not seem to matter which of the US states a participant is from. But internationally, it does seem to matter what country one is from. Beyond that conclusion, how, exactly, should we interpret this difference in response pattern similarity, between an empirically found average r of .69 across countries, and an r of .80 that would be obtained if countries did not make a difference? This is a matter yet to be resolved.

In any event, the ability of purely statistical methods to assess the comparability of measurements across cultures is fundamentally limited. While such methods as MGCFA and the much simpler approach used here provide interesting and useful information, the study of validity will, in the end, always require data from outside the measure being validated (Cronbach & Meehl, 1955). Two especially promising approaches are the use of anchoring vignettes to assess the comparability of the meaning of constructs across cultures (Mõttus et al., 2012), and using patterns of theoretically-predicted correlations with independently-assessed attributes of cultures (Mõttus et al., 2010). In other words, future research of the sort that Jüri Allik helped to pioneer and continues to conduct, will be needed as we continue to seek to understand the ways in which personality differs across cultures, and the ways in which it is the same.

[6] It is also highly statistically significant, given that the N for this correlation is 1,770, the number of non-redundant correlations being compared.

References

Allik, J., & McCrae, R. R. (2004). Toward a geography of personality traits: Patterns of profiles across 36 cultures. *Journal of Cross-Cultural Psychology*, *35*, 13–28.

Allik, J., & Realo, A. (2017). How valid are culture-level mean personality scores? In T. A. Church (Ed.), *The Praeger handbook of personality across cultures Vol 1: Trait psychology across cultures* (pp. 193–224). Santa Barbara, CA: Praeger.

Bartram, D. (2013). Scalar equivalence of OPQ32: Big Five profiles of 31 countries. *Journal of Cross-Cultural Psychology*, *44*, 61–83.

Byrne, B. M., & van de Vijver, F. J. R. (2010). Testing for measurement and structural equivalence in large-scale cross-cultural studies: Addressing the issue of nonequivalence. *International Journal of Testing*, *10*, 107–132.

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology*, *95*, 1005–1018.

Ching, C. M., Church, T. A., Katigbak, M. S., Reyes, J. A. S., Tanaka-Matsumi, J., Takaoka, S., … Ortiz, F. A. (2014). The manifestation of traits in everyday behavior and affect: A five-culture study. *Journal of Research in Personality*, *48*, 1–16.

Cronbach, L. J., & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302.

Davidov, E. Schmidt, P., & Schwartz, S.H. (2008). Bringing values back in: The adequacy of the European Social Survey to measure values in 20 countries. *Public Opinion Quarterly, 72,* 420–445.

Gurven, M., von Rueden, C., Massenkoff, M., Kaplan, H., & Lero Vie, M. (2013). How universal is the Big Five? Testing the five-factor model of personality variation among forager–farmers in the Bolivian Amazon. *Journal of Personality and Social Psychology*, *104*, 354–370.

Harzing, A.-W. (2006). Response styles in cross-national survey research a 26-country study. *International Journal of Cross Cultural Management*, *6*, 243–266.

Heine, S. J., Buchtel, E. E., & Norenzayan, A. (2008). What do cross-national comparisons of personality traits tell us? The case of conscientiousness. *Psychological Science, 19*, 309–313.

Hopwood, C. J., & Donnellan, M. B. (2010). How should the internal structure of personality inventories be evaluated? *Personality and Social Psychology Review*, 14, 332–346.

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. John (Eds.), *Handbook of personality: Theory and research* (pp. 102–138). New York: Guilford Press.

McCrae, R. R., Terracciano, A., & 78 Members of the Personality Profiles of Cultures Project (2005). Universal features of personality traits from the observer's perspective: Data from 50 cultures. *Journal of Personality and Social Psychology*, *88,* 547–561.

Mõttus, R., Allik, J., & Realo, A. (2010). An attempt to validate national mean scores of Conscientiousness: No necessarily paradoxical findings. *Journal of Research in Personality, 44,* 630–640.

Mõttus, R., Allik, J., Realo, A., Pullmann, H., Rossier, J., Zecca, G., … Tseung, C. N. (2012). Comparability of self-reported conscientiousness across 21 countries. *European Journal of Personality, 26,* 303–317.

Rentfrow, P. J., Gosling, S. D., & Potter, J. (2008). A theory of the emergence, persistence, and expression of geographic variation in psychological characteristics. *Perspectives on Psychological Science*, *3*, 339–369.

Schmitt, D. P., Allik, J., McCrae, R. R., Benet-Martínez, V., Alcalay, L., Ault, L., ... Zupancic, A.. (2007). The geographic distribution of Big Five personality traits: Patterns and profiles of human self-description across 56 nations. *Journal of Cross-Cultural Psychology*, *38*, 173–212.

Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology, 113,* 117–143.

Thalmayer, A. G., & Saucier, G. (2014). The questionnaire Big Six in 26 nations: Developing cross-culturally applicable Big Six, Big Five and Big Two inventories. *European Journal of Personality*, *28*, 482–496.

van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, *9*, 486–492.

Zecca, G., Verardi, S., Antonietti, J. P., Dahourou, D., Adjahouisso, M., Ah-Kion, J., ... Rossier, J. (2013). African cultures and the Five-Factor Model of Personality: Evidence for a specific Pan-African structure and profile? *Journal of Cross-Cultural Psychology, 44*, 684–700.

Appendix 11A Members of the International Situations Project.

Argentina:	Maite Beramendi, Universidad de Buenos Aires
Australia:	Brock Bastian, University of Melbourne
Austria:	Aljoscha Neubauer, University of Graz
Bolivia:	Diego Cortez, Universidad Católica Bolviana, La Paz
Bolivia:	Eric Roth, Universidad Católica Bolviana, La Paz
Brazil:	Ana Torres, Federal University of Paraíba
Brazil:	Daniela S. Zanini, Pontifical Catholic University of Goiás
Bulgaria:	Kristina Petkova, Bulgarian Academy of Sciences
Canada:	Jessica Tracy, University of British Columbia
Canada:	Catherine Amiot, Université du Québec à Montréal
Canada:	Mathieu Pelletier-Dumas, Université du Québec à Montréal
Chile:	Roberto González, Pontificia Universidad Católica de Chile
Chile:	Ana Rosenbluth, Universidad Adolfo Ibáñez
Chile:	Sergio Salgado, Universidad de La Frontera
China, Beijing:	Yanjun Guan, Durham University, United Kingdom
China, Shanghai:	Yu Yang, Shanghai Tech University
Colombia:	Diego Forero, Universidad Antonio Nariño, Bogotá and Universidad de Ciencias Aplicadas y Ambientales, Bogotá
Colombia:	Andrés Camargo, Universidad Antonio Nariño, Bogotá and Universidad de Ciencias Aplicadas y Ambientales, Bogotá
Crete:	Emmanouil Papastefanakis, University of Crete
Crete:	Georgios Kritsotakis, Technological Institute of Crete
Crete:	Irene Spyridaki, University of Crete
Crete:	Evangelia Fragkiadaki, Hellenic American University
Croatia:	Željko Jerneić, University of Zagreb
Czech Republic:	Martina Hřebíčková, Czech Academy of Sciences
Czech Republic:	Sylvie Graf, Czech Academy of Sciences
Denmark:	Pernille Strøbæk, University of Copenhagen
Estonia:	Anu Realo, University of Tartu

France:	Maja Becker, Université de Toulouse
France:	Christelle Maisonneuve, Univ Rennes, Rennes
Gaza (Palestine):	Sofian El-Astal, Al Azhar University-Gaza
Georgia:	Vladimer Gamsakhurdia, Ivane Javakhishvili Tbilisi State University
Germany:	Matthias Ziegler, Humboldt University
Germany:	Lars Penke, University of Goettingen & Leipniz Science Campus Primate Cognition
Germany:	John Rauthmann, Universität zu Lübeck
Hong Kong:	Emma E. Buchtel, The Education University of Hong Kong
Hong Kong:	Victoria Wai-Lan Yeung, Lingnan University
Hungary:	Ágota Kun, Budapest University of Technology and Economics
Hungary:	Peter Gadanecz, Budapest University of Technology and Economics
Hungary:	Zoltán Vass, Karoli Gaspar University of the Reformed Church in Hungary
Hungary:	Máté Smohai, Karoli Gaspar University of the Reformed Church in Hungary
India:	Abhijit Das, AMRI Institute of Neurosciences, Kolkata
India:	Anagha Lavalekar, Jnana Prabodihini's Institute of Psychology, Pune
Indonesia:	Meta Zahro Aurelia, Univeritas Ahmad Dahlan
Indonesia:	Dian Kinayung (translators), Univeritas Ahmad Dahlan
Indonesia:	Vanessa Gaffar, Universitas Pendidikan Indonesia
Indonesia:	Gavin Sullivan, Coventry University
Indonesia:	Christopher Day, Coventry University
Israel:	Eyal Rechter, Ono Academic College
Italy:	Augusto Gnisci, University of Campania
Italy:	Ida Sergi, University of Campania
Italy:	Paolo Senese, University of Campania
Italy:	Marco Perugini, University of Milan-Bicocca
Italy:	Giulio Costantini, University of Milan-Bicocca
Japan:	Asuka Komiya, Hiroshima University
Japan:	Tatsuya Sato, Ritsumeikan University
Japan:	Yuki Nakata, Ritsumeikan University
Japan:	Shizuka Kawamoto, Yamanashi University
Jordan:	Marwan Al-Zoubi, University of Jordan
Kenya:	Nicholas Owsley, Busara Center for Behavioral Economics
Kenya:	Chaning Jang, Busara Center for Behavioral Economics
Kenya:	Georgina Mburu, Busara Center for Behavioral Economics
Kenya:	Irene Ngina, Busara Center for Behavioral Economics
Latvia:	Girts Dimdins, University of Latvia
Lithuania:	Rasa Barkauskiene, Vilnius University
Lithuania:	Alfredas Laurinavicius, Vilnius University
Malaysia:	Khairul A. Mastor, Universiti Kebangsaan Malaysia
Mexico:	Elliott Kruse, EGADE Business School Monterrey
Mexico:	Nairán Ramírez-Esparza, Fundación Universidad de las Américas Puebla
Netherlands:	Jaap Denissen, Tilburg University
Netherlands:	Marcel Van Aken, University of Utrecht
New Zealand:	Ron Fischer, Victoria University of Wellington, Wellington

Nigeria:	Ike E. Onyishi, University of Nigeria, Nsukka
Nigeria:	Kalu T. Ogba, University of Nigeria, Nsukka
Norway:	Siri Leknes, University of Oslo
Norway:	Vera Waldal Holen, University of Oslo
Norway:	Ingelin Hansen, University of Oslo
Norway:	Christian Krog Tamnes, University of Oslo
Norway:	Kaia Klæva, University of Oslo
Pakistan:	Muhammad Rizwan, Government of Pakistan
Pakistan:	Rukhsana Kausar, University of the Punjab, Lahore
Pakistan:	Nashi Khan, University of the Punjab, Lahore
Peru:	Agustín Espinosa, Pontificia Universidad Católica del Peru
Philippines:	Maria Cecilia Gastardo- Conaco, University of Philippines-Diliman
Philippines:	Diwa Malaya A. Quiñones, University of Philippines-Diliman
Poland:	Piotr Szarota, Institute of Psychology of The Polish Academy of Sciences
Poland:	Paweł Izdebski, Kazimierz Wielki University
Poland:	Martyna Kotyśko, University of Warmia and Mazury
Portugal:	Joana Henriques-Calado, Universidade de Lisboa, Faculdade de Psicologia, CICPSI, Alameda da Universidade, Lisboa
Romania:	Florin Alin Sava, West University of Timisoara
Russia:	Olga Lvova, St. Petersburg State University
Russia:	Victoria Pogrebitskaya, St. Petersburg State University
Russia:	Mikhail Allakhverdov, St. Petersburg State University
Russia:	Sergey Manichev, St. Petersburg State University
Senegal:	Oumar Barry, Université Cheikh Anta Diop de Dakar-Sénégal
Serbia:	Snežana Smederevac, University of Novi Sad
Serbia:	Petar Čolović, University of Novi Sad
Serbia:	Dušanka Mitrović, University of Novi Sad
Serbia:	Milan Oljača, University of Novi Sad
Singapore:	Ryan Hong, National University of Singapore
Slovakia:	Peter Halama, Slovak Academy of Sciences
Slovenia:	Janek Musek, University of Ljubljana
South Africa:	Francois De Kock, University of Capetown
South Korea:	Gyuseog Han, Chonnam National University
South Korea:	Eunkook M. Suh, Yonsei University
South Korea:	Soyeon Choi, Yonsei University
Spain:	Luis Oceja, Universidad Autónoma de Madrid
Spain:	Sergio Villar, Universidad Autónoma de Madrid
Spain:	David Gallardo-Pujol, University of Barcelona
Sweden:	Zoltan Kekecs, Lund University
Sweden:	Nils Arlinghaus, Lund University
Sweden:	Daniel P. Johnson, Lund University
Sweden:	Alice Kathryn O'Donnell, Lund University
Switzerland:	Janina Larissa Bühler, University of Basel
Switzerland:	Clara Kulich, Université de Genève
Switzerland:	Fabio Lorenzi-Cioldi, Université de Genève
Switzerland:	Mathias Allemand, University of Zurich
Taiwan:	Yenping Chang, University of North Carolina

Taiwan:	Weifang Lin, Chulalongkorn University
Thailand:	Watcharaporn Boonyasiriwat, Chulalongkorn University
Turkey:	Adil Saribay, Boğaziçi University
Turkey:	Oya Somer, Cyprus International University
Turkey:	Pelin Karakus Akalin, Istinye University, Istanbul
Uganda:	Peter Kakubeire Baguma, Makerere University
Ukraine:	Alexander Vinogradov, Taras Shevchenko National University of Kyiv
Ukraine:	Larisa Zhuravlova, Zhytomyr Ivan Franko State University
United Kingdom:	Jason Rentfrow, University of Cambridge
United Kingdom:	Mark Conner, University of Leeds
United States, AL:	Alexa Tullett, University of Alabama
United States, CA:	Erica Baranski, University of California, Riverside
United States, CT:	Nairán Ramírez-Esparza, University of Connecticut
United States, ID:	Douglas E. Colman, Idaho State University
United States, IL:	Joey T. Cheng, University of Illinois at Urbana-Champaign
United States, TX:	Eric Stocks, University of Texas, Tyler
Viet Nam:	Huyen Thi Thu Bui, Hanoi National University of Education

12 WHY SOMETIMES A MAN IS MORE LIKE A WOMAN: INSIGHTS INTO THE "GENDER PARADOX" OF PSYCHOLOGICAL SEX DIFFERENCES AROUND THE WORLD

DAVID P. SCHMITT

Among his many achievements, Jüri Allik and his colleagues were among the first to document a cross-cultural "gender paradox" in people's self-reported personality traits. Namely, differences in how men and women describe their traits are typically larger and more conspicuous in highly gender egalitarian cultures (e.g., across Scandinavia where women and men experience more similar gender roles, sex role socialization, and sociopolitical gender equity) compared to less gender egalitarian cultures (e.g., across Africa or South/Southeast Asia). It is my honor to celebrate Jüri Allik's sterling career with this chapter on sex differences in personality traits across one of the largest number of cultures yet investigated—58 nations of the International Sexuality Description Project-2 (ISDP-2). In this dataset, the gender paradoxical findings were replicated, with sex differences in Big Five personality traits being demonstrably larger in more gender egalitarian cultures. In our current era of most findings from classic psychological science failing to replicate, this successful replication serves as a testament to Jüri Allik's status as among the most rigorous and prescient scientists within the field of personality psychology.

Introduction

Most research on psychological differences between men and women has found, for the most part, men and women are not all that different (Hyde, 2005; Zell, Krizan, & Teeter, 2015; cf. Del Giudice, Booth, & Irwing, 2012). For instance, in an extensive review of sex differences in personality traits, Hyde (2014) found men and women are very similar, with small differences on some measures (e.g., reward sensitivity, self-esteem), and slightly larger sex differences in measures of sensation seeking, as well as, measures of physical aggression and certain aspects of sexuality (e.g., attitudes toward casual sex and masturbation frequency).

Cross-culturally, however, men and women do not always display very small degrees of psychological difference. Psychological sex differences vary a lot in size across cultures, sometimes men and women are quite different (Costa, Terracciano, & McCrae, 2001; Löckenhoff et al., 2014; Lynn & Martin, 1997). Moreover, variation in the size of psychological sex differences across cultures is not random, it appears to be

E-mail: david.schmitt@brunel.ac.uk

highly patterned (Allik & McCrae, 2004; Schmitt, 2015, 2017). Strangely, one of the patterns is that psychological sex differences are larger in cultures with more gender egalitarianism (i.e., more similar gender roles, gender-equal sex role socialization, and sociopolitical gender equity; Schmitt, Realo, Voracek, & Allik, 2008). This counter-intuitive pattern is a form of the "Nordic gender paradox" and appears to be a reliable phenomenon (Gracia & Merlo, 2016).

Among the earliest studies to document that sex differences in personality are larger in cultures with more gender egalitarianism involved the assessment of Five-Factor Model or Big Five personality traits (i.e., extraversion, agreeableness, conscientiousness, neuroticism, and openness; see Costa et al., 2001; McCrae & Allik, 2002). In a study of Big Five sex differences across 55 nations, Schmitt and colleagues (2008) found that men self-reported lower levels than women did of neuroticism (overall d = -0.40), agreeableness (d = -0.15), conscientiousness (d = -0.12), and extraversion (d = -0.10) across most nations. However, the more gender egalitarian the culture, the larger were sex differences in personality traits. The largest overall sex differences in personality were found in relatively high gender egalitarian cultures of France (d = -0.44) and the Netherlands (d = -0.36), whereas the smallest sex differences were found in the relatively low gender egalitarian cultures of Botswana (d = 0.00) and India (d = -0.01). The same counter-intuitive pattern of findings—larger sex differences in the Big Five personality traits being found in more gender egalitarian cultures—had been previously documented by Costa et al. (2001) and has since been replicated across dozens of cultures by Lippa (2010) and Mac Giolla and Kajonius (2018).

In this chapter, the focus is on the issue of sex differences in personality traits varying in size across cultures, and what might explain the curious case of the "gender paradox" in personality (Schmitt et al., 2008).

Method

Participants

The data reported in this chapter comes from the International Sexuality Description Project-2 (ISDP-2), a collaborative effort in 2005 that involved over 200 social, behavioral, and biological scientists from 58 nations (N = 36,314, including 15,205 men and 21,109 women; Schmitt et al., 2017). A detailed description of the methodology and sampling techniques used in the ISDP-2 is given elsewhere (Schmitt et al., 2017).

Measuring Personality Traits

All samples in the ISDP-2 were administered the Big Five Inventory (BFI) of personality traits (Benet-Martínez & John, 1998). The 44-item English BFI was constructed to allow quick and efficient assessment of five personality dimensions—extraversion, agreeableness, conscientiousness, neuroticism, and openness—when

there is no possibility or need for more differentiated measurement of personality facets (Benet-Martínez & John, 1998). Self-report ratings are made on a scale from 1 (*disagree strongly*) to 5 (*agree strongly*) for each of the 44 items. A more detailed description of the samples and psychometric qualities of the BFI in the ISDP are given elsewhere (see Schmitt et al., 2007).

Measuring Gender Equality

In this chapter, cultural levels of gender egalitarianism were indexed in six ways. Capturing gender egalitarianism using such a diverse range of approaches is an important step at overcoming limitations of any one method of capturing cultural gender egalitarianism (Allik, 2005; AlMutairi, Yen, & Heller, 2018).

(1) *Gender Empowerment Measure* (GEM) is a composite of a culture's level of gender egalitarianism measuring three basic dimensions of empowerment—economic participation and decision-making, political participation and decision-making, and power over economic resources. For example, the GEM quantifies women's (relative to men's) estimated earned income, their percentage share of parliamentary seats, and their relative percentage as legislators, senior officials, and managers. National GEM data from the United Nations Development Programme (2007) were available for 48 nations of the ISDP-2.

(2) *Gender Gap Index* (GGI) reflects gender equality through women's (relative to men's) economic participation and opportunity, educational attainment, political empowerment, and health and survival. Hence, the GGI includes health and survival, whereas the GEM does not (see Else-Quest, Hyde, & Linn, 2010).

(3) *Gender Equality Index* (GEQ) focuses on gender equality in life expectancy, education, and per capita income indicators. Hence, the GEQ includes health and survival, but does not include political equality (see Else-Quest et al., 2010).

(4) *Standardized Index of Gender Equality* (SIGE) includes women's (relative men's) access to education, life expectancy, economic activity, labor market participation, and share of parliamentary seats. The SIGE tends to emphasize economic gender equality more than the former three indicators (see Else-Quest et al., 2010).

(5) A fifth measure of cultural gender egalitarianism comes directly from responses to questions about gender equality attitudes from the *World Values Survey* (WVS; Inglehart & Norris, 2003). The WVS included representative samples of dozens of nations, with higher scores on the WVS measure of gender equality attitudes indicating more positive attitudes toward gender equality (see Inglehart & Norris, 2003 for scaling details).

(6) A sixth measure of cultural gender egalitarianism was from responses to 10 counter-balanced questions from the *Sex-Role Ideology Scale* (SRI) from the ISDP-2 (Kalin & Tilby, 1978). Higher scores on the SRI measure indicate more progressive attitudes toward gender equality (e.g., agreeing with the statement "Married women should be able to have men as friends"). The more women and men expressive progressive attitudes toward gender equality, the greater a culture's level of gender egalitarianism.

Results and Discussion

The Gender Paradox Examined in the ISDP-2

Critical to evaluating the cross-cultural gender paradox of personality is correlating the size of a culture's sex differences in Big Five personality traits to the culture's level of gender egalitarianism. For example, women tend to report slightly higher levels of broadly measured extraversion (Feingold, 1994). In a study of gender differences in Big Five traits across 55 nations, Schmitt and his colleagues (2008) found across most nations that women reported higher levels of extraversion than men did (d = -0.10), similar to levels observed in the ISDP-2 (d = -0.14). As seen in Table 12.1, men's and women's extraversion levels in the ISDP-2 tended to be positively correlated with measures of cultural gender egalitarianism. However, these links were stronger among women around the world leading to larger sex differences being apparent in more gender egalitarian cultures—supporting evidence of a continuing gender paradox of personality and replicating the results of Schmitt and colleagues (2008). Specifically, sex differences in extraversion were negatively correlated with GEM, $r(46) = -.38$, $p < .01$, WVS, $r(32) = -.36$, $p < .05$, and SRI, $r(50) = -.32$, $p < .05$.

Schmitt and his colleagues (2008) found across most nations that women reported higher levels of agreeableness than men did (d = -0.15), similar to levels observed in the ISDP-2 (d = -0.17). As seen in Table 12.1, men's agreeableness levels in the ISDP-2 were negatively correlated with some measures of cultural gender egalitarianism. Hence, larger sex differences tended to be more apparent in more gender egalitarian cultures. For instance, sex differences in agreeableness were negatively correlated with GEM, $r(46) = -.25$, $p < .05$, WVS, $r(32) = -.38$, $p < .05$, and SRI, $r(50) = -.35$, $p < .01$. As shown in Figure 12.1, some of the largest sex differences in agreeableness (i.e., most negative d scores, reflecting women are much more agreeable than men) are found in nations with highly progressive sex role ideologies such as Iceland (d = -0.32) and Norway (d = -0.39), whereas women are not more agreeable than men in less gender egalitarian nations of Nigeria (d = 0.18) and Bangladesh (d = 0.36).

Table 12.1 Correlations between Big Five personality traits (men's, women's, and sex differences) and six gender egalitarianism indicators across nations of the International Sexuality Description Project-2.

	Gender Egalitarianism Indicators					
Big Five Trait	GEM	GGI	GEQ	SIGE	WVS	SRI
Extraversion						
Men	.25*	.20	.27	.21	.34*	.29*
Women	.43***	.32*	.36*	.26	.52***	.42**
Sex Difference (*d*)	-.38**	-.23	-.19	-.09	-.36*	-.32*
Agreeableness						
Men	.10	.17	.09	.05	-.23	-.25*
Women	.23	.17	.05	.02	-.04	-.05
Sex Difference (*d*)	-.25*	-.05	.06	.05	-.38*	-.35**
Conscientiousness						
Men	.02	.13	-.16	.03	-.09	-.15
Women	.19	.13	-.11	-.05	.19	.09
Sex Difference (*d*)	-.27*	-.01	-.03	.12	-.44**	-.36**
Neuroticism						
Men	-.27*	-.42**	-.19	-.37**	-.05	-.03
Women	-.13	-.29*	-.10	-.16	.07	.24*
Sex Difference (*d*)	-.06	-.22	-.15	-.38**	-.13	-.32*
Openness						
Men	.16	.03	.01	-.03	.08	.41***
Women	.09	-.03	.08	.06	.09	.38**
Sex Difference (*d*)	.06	.05	-.10	-.13	-.03	-.04

Note. GEM = Gender Empowerment Measure ($N = 48$); GGI = Gender Gap Index ($N = 39$); GEQ = Gender Equality Index ($N = 41$); SIGE = Standardized Index of Gender Equality ($N = 42$); WVS = World Values Survey Gender Equality Attitudes Measure ($N = 34$); SRI = Sex Role Ideology Scale (higher scores = more progressive; $N = 52$).

***$p < .001$, **$p < .01$, *$p < .05$

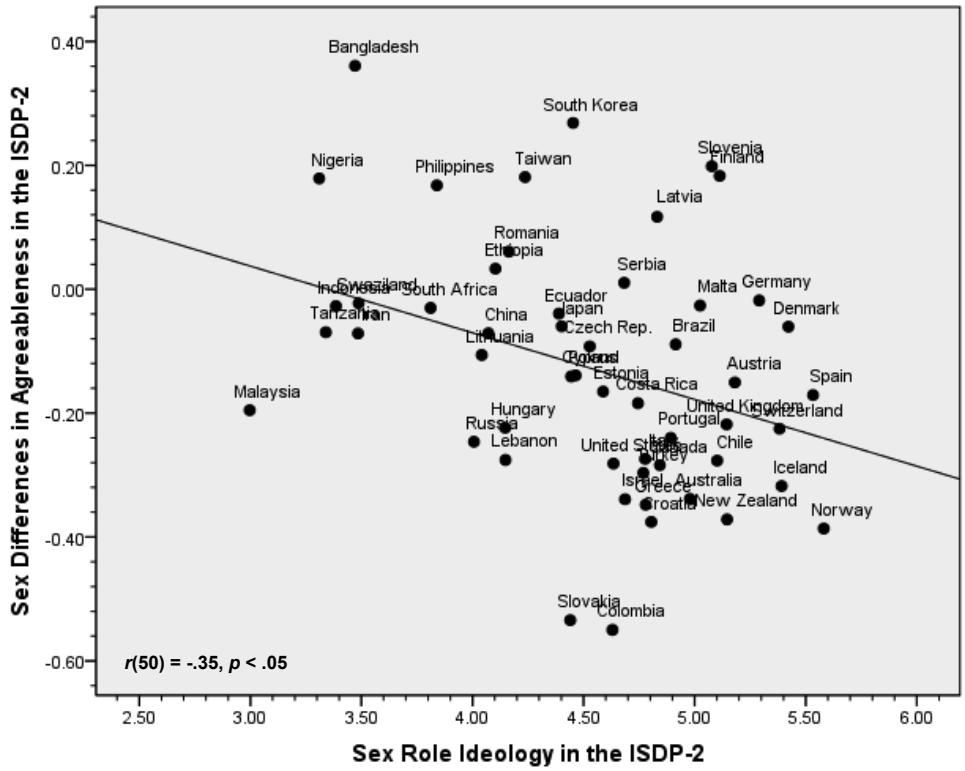


Figure 12.1 National sex differences in Agreeableness (*d*) related to national levels of the Sex Role Ideology (SRI) score across 52 nations of the International Sexuality Description Project-2 (ISDP-2). Higher scores indicate more progressive sex role ideology (x-axis) and smaller sex differences (y-axis).

Schmitt and his colleagues (2008) found women reported higher levels of neuroticism than men did (d = -0.40) across most cultures, similar to levels observed in the ISDP-2 (d = -0.49). As seen in Table 12.1, men's and women's neuroticism levels in the ISDP-2 tended to be negatively correlated with measures of cultural gender egalitarianism. However, these links were stronger among women around the world leading to larger sex differences being apparent in more gender egalitarian cultures. For instance, sex differences in neuroticism were negatively correlated with SIGE, $r(40)$ = -.38, $p < .01$ and SRI, $r(50)$ = -.32, $p < .05$. As shown in Figure 12.2, some of the largest sex differences in neuroticism (i.e., most negative d scores, reflecting women are much more neurotic than men) are found in nations with highly progressive sex role ideologies such as New Zealand (d = -0.64) and Finland (d = -0.74), whereas women are not more neurotic than men in less gender egalitarian nations of Ethiopia (d = -0.15) and Tanzania (d = -0.11).

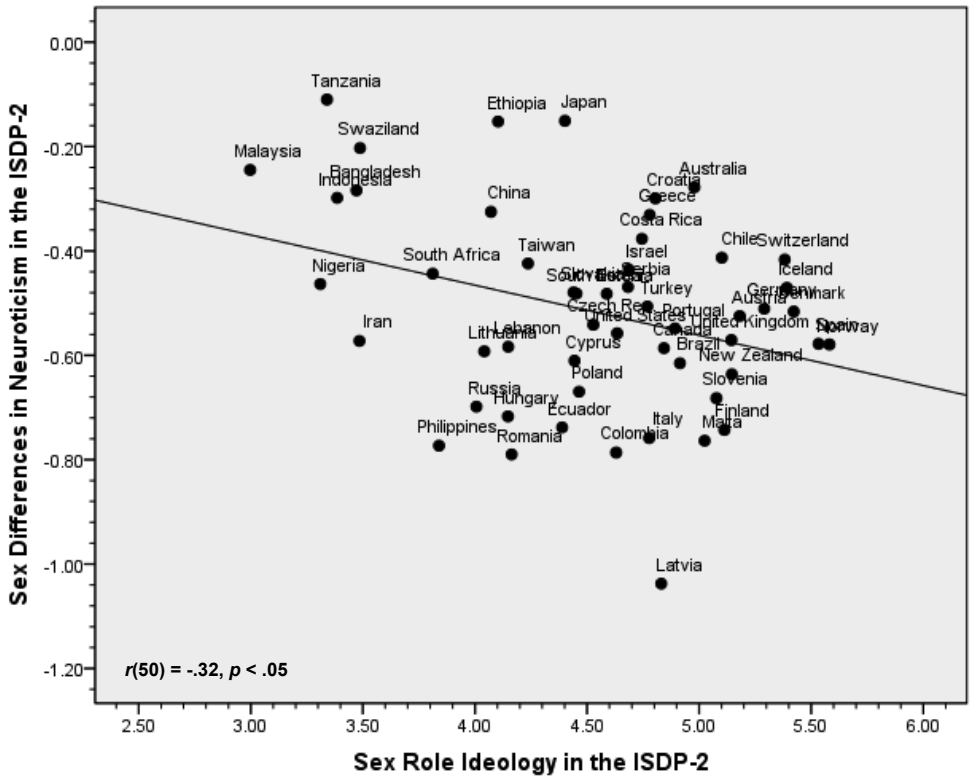


Figure 12.2 National sex differences in Neuroticism (*d*) related to national levels of the Sex Role Ideology (SRI) score across 52 nations of the International Sexuality Description Project-2 (ISDP-2). Higher scores indicate more progressive sex role ideology (x-axis) and smaller sex differences (y-axis).

Explaining the Gender Paradox of Personality across Cultures

Sex differences in many aspects of personality, sexuality, and cognition are much larger in cultures with more egalitarian sex role socialization and greater socio-political gender equity. This includes sex differences in Big Five personality traits such as extraversion, agreeableness, and neuroticism; Dark Triad traits such as Machiavellianism, narcissism, psychopathy; romantic orientations such as dismissing attachment, emotional investment, enjoying casual sex, and mate preferences for attractiveness; cognitive abilities such as spatial location ability and spatial rotation ability; personal and occupational values such as benevolence values and empathetic occupational preferences; and many other traits such as self-esteem, subjective well-being, depression, and social dominance orientation (Schmitt, 2015).

Even sex differences in physical traits such as height, obesity, and blood pressure are conspicuously larger in cultures with more cultural gender egalitarianism (Schmitt, 2015). This suggests it is unlikely that larger sex differences in personality traits are due to more traditional sex role socialization or patriarchy. Instead, evolutionary theories

involving neuroandrogenic (De Bolle, 2015; Ellis, 2011; Greenberg, Warrier, Allison, & Baron-Cohen, 2018) and socioecological factors (Low, 1998; Lukaszewski, Gurven, von Rueden, & Schmitt, 2017; Schmitt, 2014, 2015) may be better at explaining the size of psychological sex differences across cultures.

For instance, evolutionary psychologists expect sexual selection has sculpted psychological sex differences in humans, just as it has done with all other sexually-reproducing species on the planet (Buss, 1995; Mealey, 2000; Schmitt, 2017). However, it is also expected that the phenotypic emergence of evolved sex differences is sometimes designed to be attenuated or accentuated via facultatively adaptive responses to socioecological factors (e.g., local pathogens evoke women's mate preferences for physical attractiveness more than men's; Gangestad, Haselton, & Buss, 2006). Moreover, evolutionary psychologists further expect evolved sex differences to vary in size across cultures due to the moderating effects of other psychological adaptations and socioecological factors: religious suppression of sexuality, for instance, may be designed to be more restrictive of women's than men's short-term mating behavior (Schmitt & Fuller, 2015).

It is likely to take many theories—evolutionary and otherwise—in complex combinations to fully explain the "gender paradox" patterns of sex differences in personality we see across cultures. The work of Jüri Allik and his colleagues has laid the empirical, psychometric, and theoretical foundations necessary for building this explanatory lattice (Allik & McCrae, 2002; Allik, Realo, & McCrae, 2013; Schmitt et al., 2008). Personality psychology stands on this important work, and we are grateful for the enhanced view he has provided.

References

Allik, J. (2005). Personality dimensions across cultures. *Journal of Personality Disorders, 19,* 212–232.

Allik, J., & McCrae, R. R. (2002). A Five-Factor Theory perspective. In R. R. McCrae & J. Allik (Eds.), *The Five-Factor Model of personality across cultures* (pp. 303–322). New York: Kluwer Academic/Plenum Publishers.

Allik, J., & McCrae, R. R. (2004). Toward a geography of personality traits: Patterns of profiles across 36 cultures. *Journal of Cross-Cultural Psychology, 35*, 13–28.

Allik, J., Realo, A., & McCrae, R. R. (2013). Universality of the Five-Factor Model of personality. In T. A. Widiger & P. T. Costa, Jr. (Eds.), *Personality disorders and the Five-Factor Model of personality* (pp. 61–74). Washington, DC, US: American Psychological Association.

AlMutairi, S., Yen, D., & Heller, M. (2018). The many facets of national culture: A critical appraisal. *Comparative Sociology, 17*, 759–781.

Benet-Martínez, V., & John, O. P. (1998). Los Cinco Grandes across cultures and ethnic groups: Multitrait-multimethod analyses of the Big Five in Spanish and English. *Journal of Personality and Social Psychology, 75*, 729–750.

Buss, D. M. (1995). Psychological sex differences. Origins through sexual selection. *American Psychologist, 50,* 164–168.

Costa, P. T., Jr., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology, 81,* 322–331.

De Bolle, M., De Fruyt, F., McCrae, R. R., Löckenhoff, C. E., Costa, P. T., Jr., Aguilar-Vafaie, M. E., ... Terracciano, A. (2015). The emergence of sex differences in personality traits in early adolescence: A cross-sectional, cross-cultural study. *Journal of Personality and Social Psychology, 108*, 171–185.

Del Giudice, M., Booth, T., & Irwing, P. (2012). The distance between Mars and Venus: Measuring global sex differences in personality. *PLOS One, 7,* e29265. doi:10.1371/journal.pone.0029265

Ellis, L. (2011). Evolutionary neuroandrogenic theory and universal gender differences in cognition and behavior. *Sex Roles, 64,* 707–722.

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin, 136,* 103–127.

Feingold, A. (1994). Gender differences in personality: A meta-analysis. *Psychological Bulletin, 116,* 429–456.

Gangestad, S. W., Haselton, M. G., & Buss, D. M. (2006). Evolutionary foundations of cultural variation: Evoked culture and mate preferences. *Psychological Inquiry, 17,* 75–95.

Gracia, E., & Merlo, J. (2016). Intimate partner violence against women and the Nordic paradox. *Social Science & Medicine, 157*, 27–30.

Greenberg, D. M., Warrier, V., Allison, C., & Baron-Cohen, S. (2018). Testing the empathizing–systemizing theory of sex differences and the extreme male brain theory of autism in half a million people. *Proceedings of the National Academy of Sciences, 115*, 12152–12157.

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60,* 581–592.

Hyde, J. S. (2014). Gender similarities and differences. *Annual Review of Psychology, 65,* 373–398.

Inglehart, R., & Norris, P. (2003). *Rising tide: Gender equality and cultural change around the world.* Cambridge, UK: Cambridge University Press.

Kalin, R., & Tilby, P. J. (1978). Development and validation of a sex-role ideology scale. *Psychological Reports, 42,* 731–738.

Lippa, R. A. (2010). Sex differences in personality traits and gender-related occupational preferences across 53 nations: Testing evolutionary and social-environmental theories. *Archives of Sexual Behavior, 39,* 619–636.

Löckenhoff, C. E., Chan, W., McCrae, R. R., De Fruyt, F. D., Jussim, L., De Bolle, M. D, ... Terracciano, A. (2014). Gender stereotypes of personality: Universal and accurate? *Journal of Cross-Cultural Psychology, 45*, 675–694.

Low, B. S. (1998). The evolution of human life histories. In C. Crawford & D. L. Krebs (Eds.), *Handbook of evolutionary psychology* (pp. 131–161). Mahwah, NJ: Erlbaum.

Lukaszewski, A. W., Gurven, M., von Rueden, C. R., & Schmitt, D. P. (2017). What explains personality covariation? A test of the socioecological complexity hypothesis. *Social Psychological and Personality Science, 8*, 943–952.

Lynn, R., & Martin, T. (1997). Gender differences in extraversion, neuroticism, and psychoticism in 37 countries. *Journal of Social Psychology, 137,* 369–373.

Mac Giolla, E., & Kajonius, P. J. (2018). Sex differences in personality are larger in gender equal countries: Replicating and extending a surprising finding. *International Journal of Psychology*. Advance online publication, doi:10.1002/ijop.12529.

McCrae, R. R., & Allik, J. (Eds.) (2002). *The Five-Factor Model of personality across cultures*. New York: Kluwer Academic/Plenum Publishers.

Mealey, L. (2000). *Sex differences: Developmental and evolutionary strategies*. San Diego, CA: Academic Press.

Schmitt, D. P. (2014). Evaluating evidence of mate preference adaptations: How do we really know what *Homo sapiens* really want? In V. A. Weekes-Shackelford & T. K. Shackelford (Eds.), *Evolutionary perspectives on human sexual psychology and behavior* (pp. 3–39). New York: Springer.

Schmitt, D. P. (2015). The evolution of culturally-variable sex differences: Men and women are not always different, but when they are … it appears not to result from patriarchy or sex role socialization. In V. A. Weekes-Shackelford & T. K. Shackelford (Eds.), *The evolution of sexuality* (pp. 221–256). New York: Springer.

Schmitt, D. P. (2017). Sexual dials (not switches) theory: An evolutionary perspective on sex and gender complexity. *Evolutionary Studies in Imaginative Culture, 1,* 93–102.

Schmitt, D. P., Alcalay, L., Allik, J., Alves, I. C. B., Anderson, C. A., Angelini, A. L., …. Kökény, T. (2017). Narcissism and the strategic pursuit of short-term mating: Universal links across 11 world regions of the International Sexuality Description Project-2. *Psihologijske Teme [Psychological Topics], 26*, 89–137.

Schmitt, D. P., Allik, J., McCrae, R. R., Benet-Martínez, V., Alcalay, L., Ault, L., . . . Zupancic, A. (2007). The geographic distribution of Big Five personality traits: Patterns and profiles of human self-description across 56 nations. *Journal of Cross-Cultural Psychology, 38*, 173–212.

Schmitt, D. P., & Fuller, R. C. (2015). On the varieties of sexual experience: A cross-cultural exploration of the links between religiosity and human mating strategies. *Psychology of Religion and Spirituality, 7*, 314–326.

Schmitt, D. P., Realo, A., Voracek, M., & Allik, J. (2008). Why can't a man be more like a woman? Sex differences in Big Five personality traits across 55 cultures. *Journal of Personality and Social Psychology, 94*, 168–182.

United Nations Development Programme (2007). *Human development report 2007.* New York: Oxford University Press.

Zell, E., Krizan, Z., & Teeter, S. R. (2015). Evaluating gender similarities and differences using metasynthesis. *American Psychologist, 70*, 10–20.

13 PERSONALITY LOOSELY MIRRORS THE GEOGRAPHY OF THE WORLD

UKU VAINIK, LIISI KÖÖTS-AUSMEES, AND RENÉ MÕTTUS

Country differences in mean personality profiles are small but consistent. Similar findings have emerged in genetics, where the map of Europe can be recovered from genetic signals. It has been considered impossible to reliably link country personality profiles to their geographical locations. Here, we challenge this notion. We employed two country mean personality profile datasets, namely 51 NEO PI-R/3 country profiles and 76 IPIP-NEO country profiles. Scores on the two first principal components extracted from the IPIP-NEO profile set were robustly related to longitude and latitude (*r*s = .30 to .60); within the NEO PI-R/3 dataset, only the second component related to absolute latitude ($r = .60$, False Discovery Rate $p < .05$). When comparing domains and facets in how they associated with longitude and latitude, the datasets agreed on the facet-based longitude profiles and domain-based latitude profiles, allowing the latitude ($r = .39$) and longitude ($r = .50$) of countries within the IPIP-NEO datasets to be predicted using the corresponding associations trained in the NEO PI-R/3 dataset. In the reverse direction, only longitude could be predicted ($r = .55$). Such accuracy is considerably smaller than the associations between genetic variations and longitude/latitude, but some prediction is possible. Therefore, the current results highlight personality data as a potentially promising tool for mapping populations in a geographically meaningful way, but more features are likely needed for greater accuracy.

Introduction

That people from different countries vary in personality has long been a popular idea among lay-persons and scientists alike (Realo et al., 2009). Indeed, there may be robust between-country differences in the average personality profiles (e.g., Allik et al., 2017). But the "culture coefficient" is likely very small: country-differences may only account for about 2% of overall individual differences (Mõttus, Vainik, & Allik, 2018). Moreover, country-differences in personality scores modestly—at best—converge with lay perceptions of national character (Allik, Mõttus, & Realo, 2010; Hřebíčková, Mõttus, Graf, Jelínek, & Realo, 2018; McCrae, Terracciano, & Members of the Personality Profiles of Cultures Project, 2005), and are often outright counter-

Uku Vainik was supported by the Estonian Research Council Personal Post-doctoral Research Funding project PUTJD654 and by Fonds de recherche du Québec–Santé (FRQS) foreign post-doctoral training award. Liisi Kööts-Ausmees and René Mõttus were supported by institutional research funding (IUT2-13) from the Estonian Ministry of Education and Research.

E-mail: uku.vainik@gmail.com and rene.mottus@ed.ac.uk

intuitive (Mõttus, Allik, & Realo, 2010). But because the degree to which country-rankings in trait scores and trait-rankings vary across cultures is robust, this begets the question of what kind of information is hidden in country-level personality profiles. There must be some signal. For instance, it has been observed that geographically and culturally proximate regions tend to have similar personality profiles (Allik et al., 2017; Allik & McCrae, 2004). At the same time, personality-proximity does not seem to align with geographical proximity. Allik and colleagues (2017) gave several examples of how certain countries seem to group with countries that are not in their immediate vicinity. Their conclusion states that "It is obviously impossible to reproduce a geographic map of the people's habitat based on similarities between personality profiles alone" (Allik et al., 2017, p. 410).

Yet, the impossibility of predicting people's personality scores from the geographical position of their country has not been conclusively demonstrated; it was in fact only recently observed in that same particular analysis—a multidimensional scaling of average personality profiles (Allik et al., 2017). In principle, personal data pertaining to other kinds of markers of individual differences can be used to map the location of people with surprising accuracy. For instance, the map of Europe and several other regions can be well recovered from genetic data (Nelis et al., 2009; Novembre et al., 2008). It can be considered quite remarkable that the first two rotated principal components of variation in single nucleotide polymorphisms have correlations of r = .87 to .88 with the longitude and latitude of the locations where the data were collected.

Aim of the Present Study

Inspired by genetics and defying initial impressions that geographic coordinates and average personality scores have limited correspondence, the current analysis sought to take up the "impossible challenge" (Allik et al., 2017) of reproducing geographic maps from country mean-level personality data. For this purpose, we used personality instruments characterizing countries along the 30 personality facets operationalized in the NEO Personality Inventories (NEO PI-R, Costa & McCrae, 1992; NEO PI-3, McCrae, Costa, & Martin, 2005), as well as the International Personality Item Pool version (IPIP-NEO, Goldberg, 1999; Johnson, 2014). While we do not seek to replicate the impressive accuracy of genetic localization, due to having less data points (hundreds of thousands of polymorphisms vs. 30 personality facets), we tested whether there was *any* signal in personality scores with which we could predict the longitude and latitude of a sample from which the scores came. We first tried the principal components approach, which has been the workhorse in genetics. Unlike genetics, we extracted the two principal components from the country-level means and correlated these with the geographical coordinates of the countries. In addition, we calculated the mean personality profiles of longitude and latitude to see if the resultant models could predict the longitude and latitude of the countries in a different dataset. Such personality profile-based prediction is a fruitful approach to predicting individual-

level phenotypes, such as psychopathy (Benning, Patrick, Blonigen, Hicks, & Iacono, 2005) and body mass index (Vainik, Dagher, et al., 2018).

We relied on two different sets of country-samples tested with two conceptually similar but different instruments: NEO PI-R/3 and IPIP-NEO. Notably, the latter set of country-samples had been tested in only one language, English. This is important because it removes one of the potentially confounding factors from country-comparisons – possible incomparabilities resulting from instrument translations.

Methods

Personality Measures

We relied on the Five-Factor Model of personality, as conceptualized by the NEO PI-R/3 questionnaire (Costa & McCrae, 1992; McCrae et al., 2005). These questionnaires measure personality as five broad domains, each consisting of 6 facets. The proprietary NEO PI-R/3 measures personality dimensions with 240 questions, however the same facets can also be measured with open source alternatives, such as the IPIP-NEO versions of the NEO PI-R questionnaire (Goldberg, 1999; Johnson, 2014).

We used previously collected data. One data source was Allik and colleagues (2017), which collated NEO PI-R/3 scores from published and unpublished sources, totaling 71,870 participants from 76 samples and 62 different countries or cultures and 37 different languages. As the geographical data sources that we used (see below) did not provide region-specific coordinates (e.g., German-speaking and French-speaking parts of Switzerland) or the coordinates for historic countries (e.g., Yugoslavia), we excluded personality profiles pertaining to within-country regions or historic countries from Allik and colleagues (2017). We also selected only the largest country-sample in cases where more than one was available. After these exclusions, all analyses were based on data from 48,590 individuals representing 51 countries and 33 languages. The mean ages of samples were not documented in the source.

To test the robustness of any findings, we leveraged another dataset measuring the same domains and facets using a different questionnaire, IPIP-NEO; in these data, all participants completed the instrument in English (Johnson, 2014). From an initial sample of 926,463 participants who completed the 120-item IPIP-NEO (307,313 people responded to the items within a large test of 300 items) from altogether 234 countries, we focused on countries that had at least 200 respondents, including 100 men and 100 women. The original sample was collected through the author's personal website[1] as an effort to validate the 120-item version of IPIP-NEO (see Johnson, 2014, for further details). All analyses were conducted on country mean profiles across 15,200 participants from 76 countries. Mean ages of country-samples varied from 21.41 to 33.03 years (median 27.16, inter-quartile range 25.90, 28.50).

[1] http://www.personal.psu.edu/~j5j/IPIP/IPIP-NEOstart.html

We used public data sources (Google Developers, 2012; MaxMind, 2019) for assigning longitude, latitude, and continent for each country. The longitude and latitude pointed to the centre of each country.

Analysis

We first sought to conceptually replicate attempts to recover geographic maps from genomic data (Novembre et al., 2008). We extracted the first two unrotated principal components from the NEO PI-R/3 and IPIP-NEO country mean facet profiles and Procrustes-rotated these towards the longitude and latitude of the countries. We also tried absolute values of longitude and latitude, to assess if distance from equator (e.g., differences in average temperature) or distance from the Greenwich meridian (e.g., differences from Western Europe) could explain our findings.

Thereafter, we sought to create the personality profiles of longitude and latitude. To do this, we correlated domains and facets with longitude and latitude. Consistency between datasets was tested with intra-class correlation (ICC) estimates and their 95% confidence intervals, using the ICC() function from psych package (Revelle, 2014), based on a mean-rating (k = 2), absolute agreement, 2-way random-effects model (ICC (2,k)). Random-effects absolute agreement models should generalize to any other similar personality study (Koo & Li, 2016).

To test the unbiased predictive power of these profiles (rather than rely on models overfit to the data at hand), the correlation profile obtained from the IPIP-NEO-based samples was applied to the NEO PI-R/3-based samples, and vice versa. For example, we used the effect sizes for NEO PI-R/3 a) domains or b) facets as weights, with which each country's average IPIP-NEO a) domain and b) facet *z*-scores were multiplied; and vice versa. The resultant weighted a) domain and b) facet scores were then averaged, yielding either domain-based or facet-based "polytrait risk scores" for longitude and latitude. This approach has been previously used in personality research (e.g., Benning et al., 2005; Mõttus, Realo, Vainik, Allik, & Esko, 2017; Vainik, Dagher, et al., 2018) and is conceptually similar to the polygenic risk scoring approach, widely used in genomic research (Dudbridge, 2013; Lee et al., 2018). We included all NEO PI-R/3 domains in the domain-based risk score and all facets in the facet-based risk score, regardless of the magnitude or significance of their respective weights, as many papers working on polygenic risk scores have shown that the best prediction can generally be achieved when all predictors are included in risk scores, no matter the effect size (Lee et al., 2018; Vainik, Baker, et al., 2018). The country-level longitude and latitude values were then correlated with the polytrait risk scores (all expressed in *z*-scores).

All analysis was conducted in Microsoft R Open 3.5.1 (R Core Team, 2013), with August 01, 2018 version of packages tidyverse, cowplot, and psych (Revelle, 2014; Wickham & RStudio, 2017; Wilke & Wickham, 2016).

Results

Unlike what Allik and colleagues (2017) predicted, the two principal components associated robustly with longitude and latitude (Figure 13.1, left) in the IPIP-NEO dataset. Similar effect sizes were also seen in the NEO PI-R/3 samples, but we were underpowered to find effects in longitude, likely due to the smaller number of countries (Figure 13.1, right). Intriguingly, the associations with latitude improved considerably when absolute latitude value (i.e., distance from equator) was used. The absolute longitude (i.e., distance from the Greenwich meridian) did not increase the association with longitude. This makes sense, because distance from equator indicates something potentially more meaningful for personality (e.g., climate; Kööts, Realo, & Allik, 2011) than distance from Europe in an east-west direction. The IPIP-NEO solution is plotted in Figure 13.2. While some concordance with the world map can be seen, inconsistencies remain.

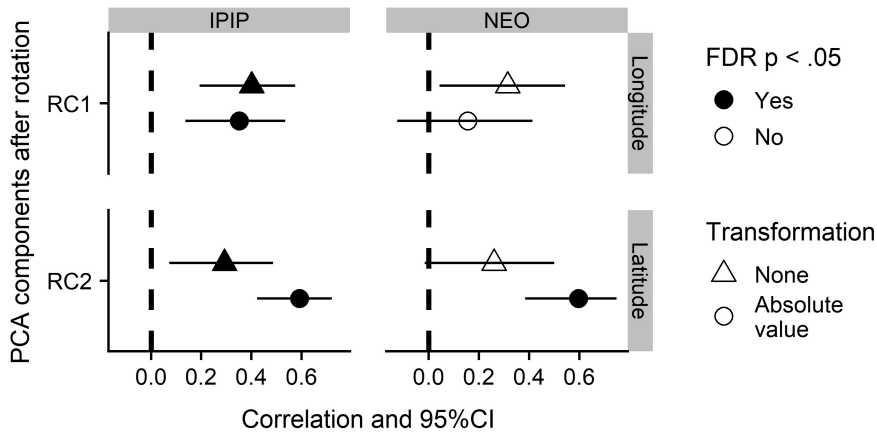


Figure 13.1 Correlations between the two Procrustes-rotated principal components (RC1 and RC2) and longitude and latitude, based on NEO PI-R/3 (N = 51 countries) and IPIP-NEO facets (N = 76 countries). CI = Confidence Interval; FDR = False Discovery Rate. Transformation = whether longitude/latitude used absolute values.

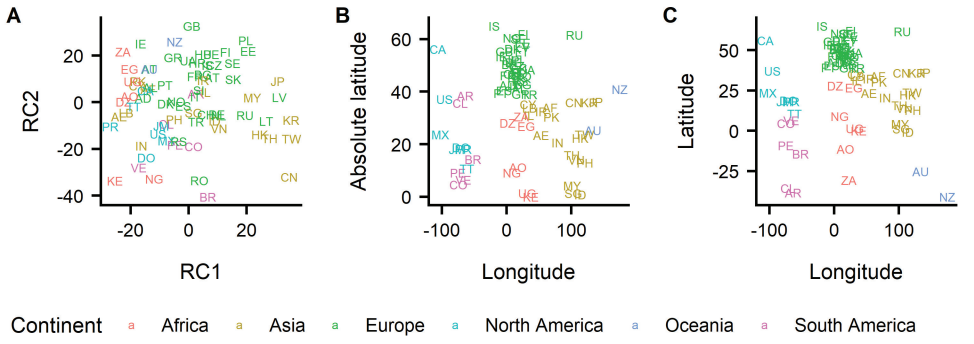


Figure 13.2 Principal component analysis (PCA) reconstruction of the world map from the IPIP-NEO personality data (N = 76 countries). (A) World map according to two Procrustes-rotated principal components (RC1, RC2) based on the IPIP-NEO dataset. The depicted solution matches better with the (B) World map, where latitude has absolute values, i.e., covers distance from equator, rather than with the (C) Actual world map.

We then tested whether the links between personality and longitude and latitude (both raw and absolute) were robust across datasets. We first found that personality domains and facets are associated with longitude and latitude within IPIP-NEO and NEO PI-R/3 separately. We then tried predicting longitude and latitude in the other dataset (NEO PI-R/3 and IPIP-NEO, respectively).

Figure 13.3 summarizes the personality facet profiles, comparing datasets and the three predicted dimensions (latitude, absolute latitude, and longitude). As can be seen, the latitude profiles from two datasets are rather different, whereas longitude profiles seem more uniform across datasets. For instance, Extraversion had opposite effects in the IPIP-NEO and NEO PI-R/3 datasets for latitude and absolute latitude. The inconsistency for latitude was mirrored by ICC between the two datasets. Namely, ICC (2,2) was poor for both latitude (ICC(2,2) = .18; 95% CI [-.52, .58], $F(29,29)$ = 1.26, p = .272) and absolute latitude (ICC(2,2) = .15; 95% CI [-.73, .59], $F(29,29)$ = 1.18, p = .329), whereas it was considerably better for longitude (ICC(2,2) = .77; 95% CI [.53, .89], $F(29,29)$ = 4.67, p < .001).

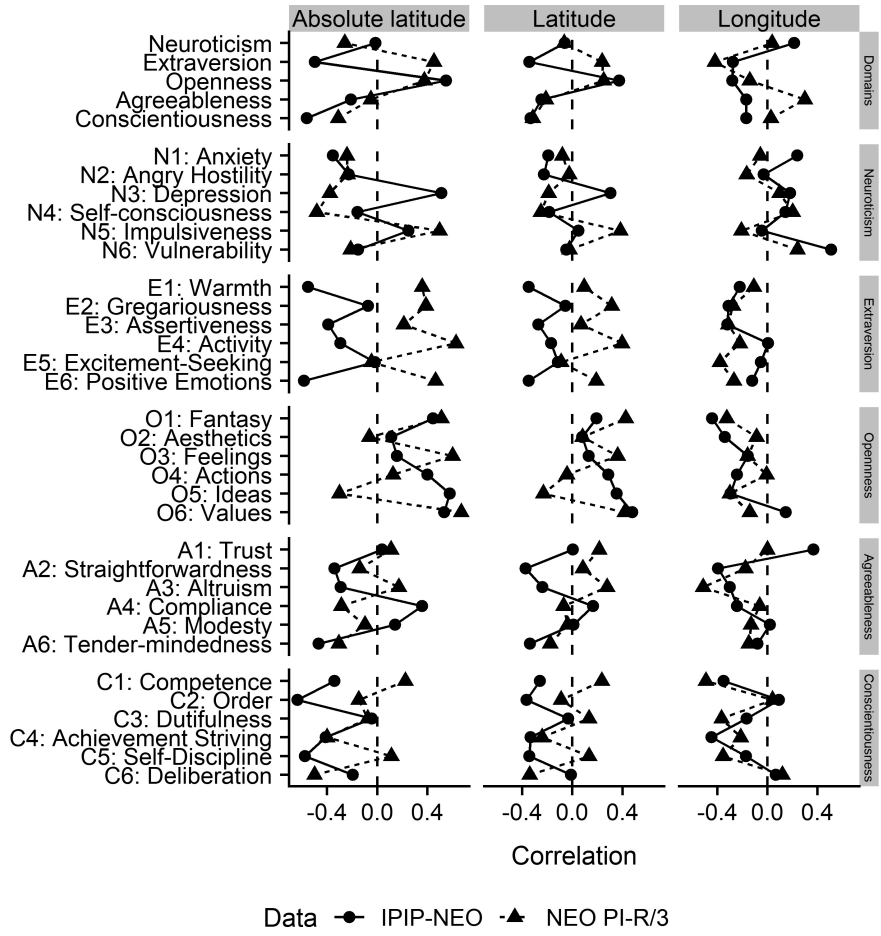


Figure 13.3 Trait—longitude/latitude correlations, by personality instrument (IPIP-NEO, *N*=76 vs. NEO PI-R/3, *N*=51).

Because of such poor findings, we repeated the analysis of between-dataset consistency for domains. While all estimates were non-significant, associations with latitude were rather consistent between the two datasets (ICC(2,2) = .70; 95% CI [-1.26, .97], $F(4,4) = 3.34$, $p = .135$), whereas the ICCs for absolute latitude (ICC(2,2) = .49; 95% CI [-3.01, .95], $F(4,4) = 1.97$, $p = .264$) and longitude (ICC(2,2) = .57; 95% CI [-2.37, .95], $F(4,4) = 2.3$, $p = .219$) were poorer.

At the same time, facet–coordinate correlations were larger for absolute latitude, which may explain the relative success of the PCA component association with absolute latitude. Namely, the mean correlation of any facet with absolute latitude was $r = .31$ /.29 for IPIP-NEO / NEO PI-R/3, whereas it was smaller for raw latitude ($r = .20$ / .18) and longitude ($r = .21$ / .20).

Such varying consistency between the two datasets was mirrored by prediction results, where personality-longitude/latitude association profiles obtained in one dataset predicted longitude/latitude in the other dataset. The profiles trained with NEO PI-R/3 facets associated well with longitude in the IPIP-NEO, and vice versa. At the same time, the success of PCA predicting (absolute) latitude was not replicated here. Further, it was the domain-based score, rather than facet-based score, that mapped to latitude in the IPIP-NEO sample. This suggests that traits predicting longitude are shared between the NEO PI-R/3 and IPIP-NEO datasets, whereas the latitude differences that PCA picks up seem to be specific to each dataset and/or instrument. Generally, the effect sizes in the NEO PI-R/3 sample were analogous to the ones found in IPIP-NEO sample, suggesting that lack of power prevented us from recovering the effects in the NEO PI-R/3 sample.

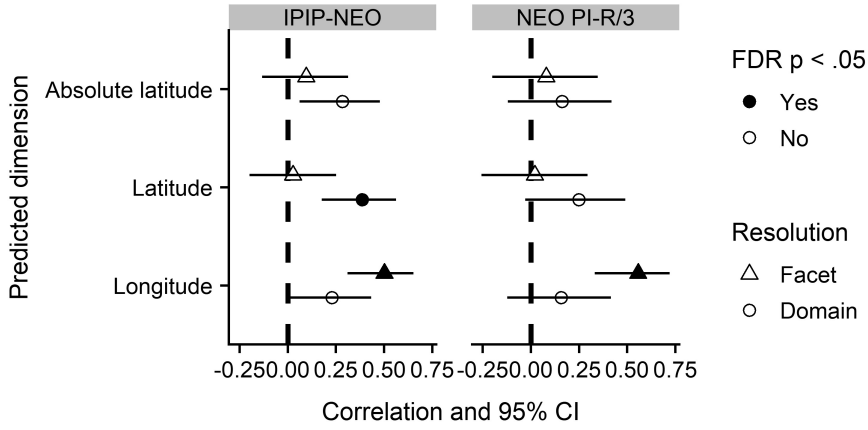


Figure 13.4 Longitude and latitude as predicted by personality risk scores trained in the other sample. CI = Confidence Interval; FDR = False Discovery Rate. Resolution = facet or domain-level analyses. IPIP-NEO, *N*=76; NEO PI-R/3, *N*=51.

Finally, we reconstructed the map of the world according to the best predictive personality scores within the IPIP-NEO dataset (Figure 13.5). Scores were chosen based on Figure 13.4—domain-score was used for latitude and facet-score for longitude. While the geographical map of the world is recognizable, there were still both consistencies and inconsistencies (Figure 13.5).

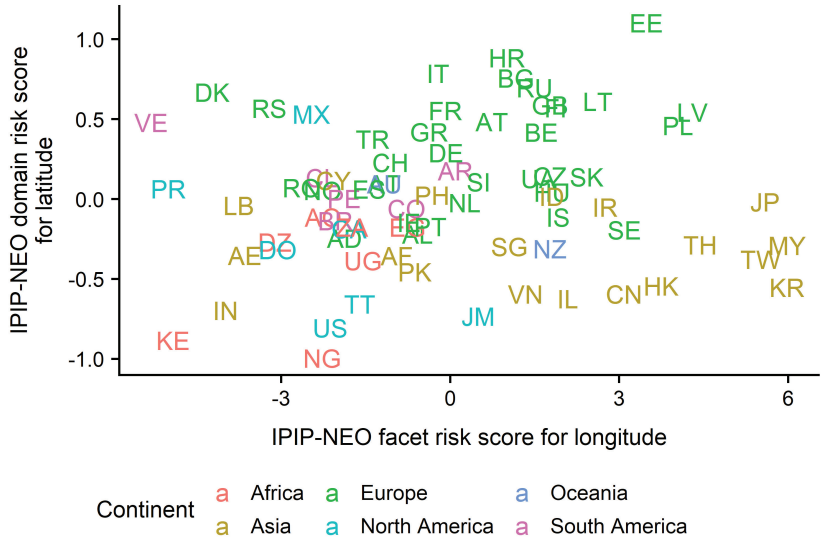


Figure 13.5 Map of the world according to the best personality risk scores applied to the IPIP-NEO dataset.

Discussion

In the current paper, we showed that personality data can be, in principle, used to predict the longitude and latitude of a country, although the predictive accuracy is moderate ($r \sim .20$ to $.60$). The accuracy is nowhere near the localization accuracy of genetic data, with a maximum correlation of $r = .60$ witnessed in the present study vs. the $r = .84$ to $.85$ that genetic PCAs have with longitude and latitude (Novembre et al., 2008). That said, the task may not be impossible, as suggested by Allik and colleagues (2017), considering that the dimensionality of the data we could work with is of a magnitude smaller than of genomic data, significantly decreasing the statistical power to detect a signal.

We succeeded in the main task—within the IPIP-NEO dataset, the first two principal components of country personality profiles aligned with longitude and latitude. The same trend appeared to be present for the NEO PI-R/3 dataset, but did not reach significance. Intriguingly, absolute latitude aligned better with the second PCA component, suggesting that personality data may reflect distance from the equator or temperature differences, rather than latitude *per se*.

But were the patterns of associations between personality traits and geographical dimensions robust across sets of samples and instruments? Even impressively strong associations may, in fact, capitalize on chance—that is, reflect overfitting (Yarkoni & Westfall, 2017). When inspecting the individual traits associated with longitude and latitude, the latitude-related traits appeared to be more sample/instrument-specific.

Therefore, while principal components can be extracted and related to personality data, their exact composition may be idiosyncratic. Strikingly, Extraversion had opposite associations in the two data sources, whereas most other inconsistencies were facet-specific. At the same time, the facet-based personality profile for longitude obtained in NEO PI-R/3 could be used to predict longitude in the IPIP-NEO sample, and vice versa. That is, personality–longitude associations may be more robust and replicable than trait–(absolute) latitude associations.

What about Conscientiousness, the trait that has arguably posed the biggest challenge in cross-cultural comparisons, with country-rankings robustly defying lay perceptions of national character as well as researchers' own intuitions (Heine, Buchtel, & Norenzayan, 2008; Mõttus et al., 2010, 2012)? First, its associations with longitude replicated well across instruments (Figure 13.3). Second, its associations with longitude appeared facet-specific, with an overall trend of countries more in the East—especially in East Asia—scoring lower (Schmitt, Allik, McCrae, & Benet-Martínez, 2007). This trend is present for the Competence (C1), Dutifulness (C3), Achievement-striving (C4), and Self-discipline (C5) facets, but not for Order (C2) or Deliberation (C6) facets; such facet-specificity is consistent with previous observations (Mõttus et al., 2010).

Strictly speaking, any result found in IPIP-NEO-based data is more likely to be attributable to actual personality differences between countries than country-differences arising from the NEO PI-R/3-based data. This is because differences between countries obtained in the NEO PI-R/3-based data could have been confounded by secondary factors, such as the language the instrument was completed in (discussed in Mõttus et al., 2018). It was therefore more pleasing to find that personality–longitude/latitude associations, either PCA-based or risk-score based, seemed more robust in IPIP-NEO than in NEO PI-R/3.

When looking at the personality maps constructed from the principal components or risk scores, more work needs to be done to better recover the world map. The maps have suggestive similarities: for instance, the continents are in roughly the right area. At the same time, huge inconsistencies remain. Possibly, greater resolution personality data, such as item-level data from the 240 NEO PI-R/3 questions could improve prediction. Such data have to be collected from larger samples to offset loss in measurement accuracy.

Taken together, the results of the current report show that personality data can do more than people think. The current findings join other recent findings that bolster the claim that personality should be taken as seriously as genetics. For instance, personality risk scores for obesity predict as much obesity in new data as polygenic risk scores do (R^2=2.3%, Vainik, Dagher, et al., 2018). Hopefully, cross-cultural personality researchers will soon be able to construct as impressive plots as population geneticists do.

References

Allik, J., Church, A. T., Ortiz, F. A., Rossier, J., Hřebíčková, M., De Fruyt, F., … McCrae, R. R. (2017). Mean profiles of the NEO Personality Inventory. *Journal of Cross-Cultural Psychology*, *48*, 402–420.

Allik, J., & McCrae, R. R. (2004). Toward a geography of personality traits: Patterns of profiles across 36 cultures. *Journal of Cross-Cultural Psychology*, *35*, 13–28.

Allik, J., Mõttus, R., & Realo, A. (2010). Does national character reflect mean personality traits when both are measured by the same instrument? *Journal of Research in Personality*, *44*, 62–69.

Benning, S. D., Patrick, C. J., Blonigen, D. M., Hicks, B. M., & Iacono, W. G. (2005). Estimating facets of psychopathy from normal personality traits: A step toward community epidemiological investigations. *Assessment*, *12*, 3–18.

Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEOFFI): Professional manual.* Odessa, FL: Psychological Assessment Resources, Inc.

Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. *PLOS Genetics*, *9*, e1003348. doi:10.1371/journal.pgen.1003348

Goldberg, L. (1999). A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe, Vol. 7* (pp. 7–28). Tilburg: Tilburg University Press.

Google Developers. (2012). *countries.csv | Dataset Publishing Language.* Retrieved January 3, 2019, from https://developers.google.com/public-data/docs/canonical/countries_csv

Heine, S. J., Buchtel, E. E., & Norenzayan, A. (2008). What do cross-national comparisons of personality traits tell us? The case of conscientiousness. *Psychological Science*, *19*, 309–313.

Hřebíčková, M., Mõttus, R., Graf, S., Jelínek, M., & Realo, A. (2018). How accurate are national stereotypes? A test of different methodological approaches. *European Journal of Personality*, *32*, 87–99.

Johnson, J. A. (2014). Measuring thirty facets of the Five Factor Model with a 120-item public domain inventory: Development of the IPIP-NEO-120. *Journal of Research in Personality*, *51*, 78–89.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*, 155–163.

Kööts, L., Realo, A., & Allik, J. (2011). The influence of the weather on affective experience. *Journal of Individual Differences*, *32*, 74–84.

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., … Cesarini, D. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics*, *50*, 1112–1121.

MaxMind. (2019). *ISO 3166 Country codes with associated continent « MaxMind Developer Site.* Retrieved January 3, 2019 from https://dev.maxmind.com/geoip/legacy/codes/country_continent/

McCrae, R. R., Costa, P. T., Jr., & Martin, T. A. (2005). The NEO PI–3: A more readable Revised NEO Personality Inventory. *Journal of Personality Assessment*, *84*, 261–270.

McCrae, R. R., Terracciano, A., & Members of the Personality Profiles of Cultures Project. (2005). Universal features of personality traits from the observer's perspective: Data from 50 cultures. *Journal of Personality and Social Psychology*, *88*, 547–561.

Mõttus, R., Allik, J., & Realo, A. (2010). An attempt to validate national mean scores of Conscientiousness: No necessarily paradoxical findings. *Journal of Research in Personality*, *44*, 630–640.

Mõttus, R., Allik, J., Realo, A., Pullmann, H., Rossier, J., Zecca, G., … Tseung, C. N. (2012). Comparability of self-reported Conscientiousness across 21 countries. *European Journal of Personality*, *26*, 303–317.

Mõttus, R., Realo, A., Vainik, U., Allik, J., & Esko, T. (2017). Educational attainment and personality are genetically intertwined. *Psychological Science*, *28,* 1631–1639.

Mõttus, R., Vainik, U., & Allik, J. (2018). The Culture Coefficient: Cross-cultural personality differences are reliable but very small. *Manuscript in preparation*.

Nelis, M., Esko, T., Mägi, R., Zimprich, F., Zimprich, A., Toncheva, D., … Metspalu, A. (2009). Genetic structure of Europeans: A view from the North-East. *PLOS One*, *4*, e5472. doi:10.1371/journal.pone.0005472

Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A. R., Auton, A., … Bustamante, C. D. (2008). Genes mirror geography within Europe. *Nature*, *456*(7218), 98–101.

R Core Team. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org/

Realo, A., Allik, J., Lönnqvist, J.-E., Verkasalo, M., Kwiatkowska, A., Kööts, L., … Renge, V. (2009). Mechanisms of the national character stereotype: How people in six neighbouring countries of Russia describe themselves and the typical Russian. *European Journal of Personality*, *23*, 229–249.

Revelle, W. (2014). *psych: Procedures for psychological, psychometric, and personality research* (Version 1.4.8.11). Retrieved December 5, 2018 from http://cran.r-project.org/web/packages/psych/index.html

Schmitt, D. P., Allik, J., McCrae, R. R., & Benet-Martínez, V. (2007). The geographic distribution of Big Five personality traits: Patterns and profiles of human self-description across 56 nations. *Journal of Cross-Cultural Psychology*, *38*, 173–212.

Vainik, U., Baker, T. E., Dadar, M., Zeighami, Y., Michaud, A., Zhang, Y., … Dagher, A. (2018). Neurobehavioral correlates of obesity are largely heritable. *Proceedings of the National Academy of Sciences*, *115*, 9312–9317.

Vainik, U., Dagher, A., Realo, A., Colodro-Conde, L., Mortensen, E. L., Jang, K., … Mõttus, R. (2018). Personality-health associations are driven by narrow traits: A meta-analysis for Body Mass Index. doi:10.31234/osf.io/z35vn

Wickham, H., & RStudio. (2017). *tidyverse: Easily install and load the "tidyverse"* (Version 1.2.1). Retrieved December 5, 2018 from https://CRAN.R-project.org/package=tidyverse

Wilke, C. O., & Wickham, H. (2016). *cowplot: Streamlined plot theme and plot annotations for "ggplot2"* (Version 0.7.0). Retrieved December 5, 2018 from https://cran.r-project.org/web/packages/cowplot/index.html

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *12*, 1100–1122.

14 FATALISM AND EXTERNAL LOCUS OF CONTROL ARE DIFFERENT CONSTRUCTS, OR *ALLIK ON ALLIKAS* (ALLIK IS THE SOURCE) OF MANY GOOD IDEAS

LAWRENCE T. WHITE AND DANIELLE R. BLAZEK

Many researchers have conceptualized fatalism as synonymous with external locus of control (LOC) and used LOC scales to measure fatalism. These practices are based on a logical error: fatalism and external LOC cannot be equivalent constructs because an attribution to fate is only one kind of external attribution. In Study 1, we developed a new, reliable measure of fatalism. In Study 2, participants completed a set of questionnaires designed to assess the degree to which fatalism and external LOC are similar or different constructs. As expected, fatalism and LOC scores correlated weakly with each other and displayed divergent patterns in terms of their correlations with other pertinent variables (e.g., political party affiliation, religiosity, happiness, and self-reported protective behaviors).

Prologue

Most people are probably unaware of Jüri Allik's valuable contributions to a unique study abroad program and to an innovative summer research program. Both programs were established by Beloit College (USA) in collaboration with the University of Tartu at a time when Jüri led the university's psychology department.

Every other year between 2000 and 2010, a group of American students came to Tartu in late August. They lived with local families and took courses in Cultural Psychology, Cross-Cultural Research Methods, Estonian Society and Identity, and Estonian Language. After 8 weeks, the students and their faculty director moved to Morocco, where they devoted another 8 weeks to a similar program at the Arabic Language Institute in Fez. Working in teams of two, students tested hypotheses by collecting data in both countries.

To our knowledge, the Beloit-Tartu-Fez program was unique. No other study abroad program allowed students to study cultural psychology, conduct cross-cultural research, and live with local families in two different countries. One of us (LTW) directed the program on four occasions, and one of us (DRB) participated as a student in the 2008 program. As department head, Jüri strongly supported the creation of the

E-mail: WhiteLT@beloit.edu and danielleblazek@gmail.com

program and handled many of the logistics. He also hired a graduate student to recruit host families, provide research support, and plan weekend excursions.

In 2005, Jüri and Anu Realo joined with LTW to direct an International Research Experience for Undergraduates (INTREU) program that was funded by the National Research Council (USA). Psychology students from different colleges and universities in the U.S. came to Tartu for 5 weeks in the summer of 2005 to collaborate with researchers in Jüri's department. The summer program produced a prize-winning conference presentation (Koller, 2006) and two research reports in the *Journal of Cross-Cultural Psychology*—one by Realo, Allik, and Greenfield (2008) that examined the relationship between social capital and individualism-collectivism and another by White, Valk, and Dialmy (2011) that compared standards of punctuality in Estonia, Morocco, and the U.S.

In the autumn of 2009, DRB asked LTW to supervise her honors thesis. DRB's experiences in Tartu and Fez had made her curious about the psychology of fatalism. As she began to read the literature, she quickly discovered that there was no consensus among researchers about how to define or measure fatalism. As LTW listened to DRB's assessment of the literature, he recalled a conversation with Jüri in the summer of 2005. Jüri had mentioned that psychological researchers sometimes invent new constructs and measures that are redundant with other existing constructs and measures. As an example, Jüri opined that one of Zimbardo's time orientations—an orientation toward the Future—was probably unnecessary because the scale designed to measure Future Orientation actually measured the personality trait of Conscientiousness. Jüri's observation inspired Maggie Koller, one of the INTREU students, to conduct the study that won a prize at the Midwestern Psychological Association meeting in 2006.

Jüri's astute observation—that researchers must take care to define and measure psychological constructs so that they are distinct and do not overlap with other constructs—served as an inspiration for our study and its title, "Fatalism and External Locus of Control Are Different Constructs."

Introduction

Researchers have conceptualized fatalism as a belief, personality trait, or attributional style that is essentially synonymous with external locus of control (e.g., Boone, de Brabandeer, Gerits, & Willeme, 1990; Cohen & Nisbett, 1998; Neff & Hoppe, 1993). Wade (1996), for example, used the terms fatalism and locus of control interchangeably, as did Goodwin and Allen (2000). McClure, Allen, and Walkey (2001) defined earthquake fatalism as attributions of earthquake damage to uncontrollable causes. Caplan and Schooler (2003) defined fatalism as "the belief that the events of one's life are largely beyond one's control" (p. 552).

Researchers who equate fatalistic thinking with an external locus of control often measure fatalism with a scale designed to assess locus of control (LOC). Cole, Rodriguez, and Cole (1978) used Rotter's (1966) Internal-External Locus of Control (I-E) Scale. Ross, Mirowsky, and Cockerham (1983) used a modified version of Rotter's I-E Scale. Roberts, Chen, and Ger (2000) used Pearlin and Schooler's (1978) mastery scale, which measures the extent to which a person regards their life-chances as being under their own control.

In short, many researchers appear to believe that fatalism and external LOC are highly similar, if not equivalent, constructs. This belief, however, is based on a logical error. Imagine a driver whose car collides with another car. How does the driver explain what happened? If the driver has an external LOC, he or she attributes the collision to external factors. Fate is one such factor, but only one of many. External factors also include other actors (e.g., the inattention of another driver), the situation (e.g., a poorly lit intersection), and luck. Fatalism is a *kind* of external LOC, but not all external causes involve fate. In short, a person can have an external LOC yet not be fatalistic because the two are theoretically separate constructs.

In our view, fatalism is a general belief that events are predetermined and inevitable. This belief can be related to other specific beliefs (about self-efficacy and the existence of a higher power, for instance), but fatalism is an overarching belief that encompasses particular beliefs without being synonymous with any one of them. We believe this definition accords closely with a lay understanding of fatalism. That is, fatalism is a broad belief about how the world works, not a specific attitude or personality trait.

In the present investigation, we sought to demonstrate empirically that fatalism and external LOC are different constructs. To that end, we constructed a new measure of fatalism and used it to investigate the degree to which fatalism and external LOC operate in similar or dissimilar fashion.

Study 1: Development of the Beloit Fatalism Scale

To our knowledge, no one has developed a context-free measure of fatalistic thinking in which fatalism is conceptualized as "a general belief that events are predetermined and inevitable." Several fatalism scales can be found in the literature, but their test-retest reliability is unknown, and they define fatalism narrowly within a specific domain, e.g., cancer fatalism (Powe, Daniels, & Finnie, 2005). The purpose of Study 1 was to develop a measure of fatalism that (a) is internally consistent, (b) can be used in a variety of circumstances, and (c) produces a wide range of scores (so as to measure individual differences precisely). The study was approved by Beloit College's institutional review board.

Participants and Method

We derived an initial pool of 26 statements by gathering items from other fatalism scales and creating new items. All statements addressed general (as opposed to domain-specific) beliefs and were stated neutrally in terms of social desirability. Statements included "I believe in fate and destiny" and "I have often found that what is going to happen will happen." Participants were asked to rate each statement on a 5-point Likert scale (1 = *strongly disagree* and 5 = *strongly agree*). In the spring of 2010, 52 students (aged 18 to 22 years) at a liberal arts college in the United States responded to all 26 items in exchange for a chance to win a $20 lottery prize.

Results

To produce a useful, internally consistent scale of fatalism, we dropped an item if participants' responses to the item showed very low variability, if the item was weakly correlated ($r < |.35|$) with all other items, or if the item's readability was poor (i.e., Flesch-Kincaid reading grade level > 10.0). Following these criteria, eight items were dropped from further consideration, resulting in an 18-item scale (see Appendix 14A) with good internal reliability (Cronbach's $\alpha = .84$). The Flesch-Kincaid readability of the Beloit Fatalism Scale (BFS) is 5.1, which means the scale can be read and understood by someone with a fifth-grade education.

We calculated a total BFS score for each participant by summing the 18 items, after reverse scoring. Scores on the BFS can range from 18 to 90. In our sample of 52 college students, scores were normally distributed and ranged from 27 to 72 ($Mdn = 43$, $M = 44.6$, $SD = 9.5$).

Study 2: Comparison of Fatalism and External Locus of Control

The purpose of Study 2[1] was to (1) assess the test-retest reliability of the BFS, (2) investigate the degree to which the BFS and Rotter's I-E Scale (1966)[2] measure similar constructs, and (3) determine the factor structure of the BFS. The study was approved by Beloit College's institutional review board.

Participants

Participants were recruited via snowball sampling in the spring of 2010. Individuals received an e-mail invitation to complete an on-line questionnaire. They were encouraged to forward the invitation to other potential participants and invited to complete a follow-up survey within 2 weeks. Each participant was eligible to win one

---

[1] DRB presented this study at the 23rd Annual Convention of the Association for Psychological Science in Washington, DC, in May 2011.

[2] We used Rotter's (1966) I-E Scale because it is the most popular measure of LOC.

of six $20 lottery prizes. Individuals who participated in Study 1 were not eligible to participate in Study 2.

Three hundred forty-three adults (228 women, 111 men, 4 unknown) completed all portions of the questionnaire. Participants ranged in age from 18 to 77 years (M = 35.3, Mdn = 26.0, SD = 16.2). Most participants (90%) identified themselves as White or Caucasian and grew up in the Midwest (48%) or Northeast (20%) regions of the U.S. Most participants described themselves as middle class (74%) and as a graduate of a 4-year college or university (81%). One hundred and nineteen participants completed the BFS a second time, 10–12 days later.

Measures and Procedure

Participants completed on-line versions of the BFS and Rotter's I-E Scale (1966), a 23-item instrument that measures the degree to which a person possesses an internal or external locus of control. A high score on the I-E Scale indicates an external locus of control.

Participants also completed a 5-item religiosity scale (White et al., 2011) and the 4-item Subjective Happiness Scale, a reliable and valid measure of global well-being (Lyubomirsky & Lepper, 1999; Matte & Schaefer, 2004). Higher scores indicate higher levels of religiosity and well-being, respectively.

Participants also reported how often they eat fast food, wear a helmet while bicycling, visit a doctor for a routine checkup, and other behaviors. In our judgment, these behaviors may reflect different levels of fatalistic thinking. Finally, participants provided information about their age, gender, education, social class, ethnicity, and political party affiliation.

Results and Discussion

BFS scores were distributed normally (with a flattened peak) and ranged from 22 to 72 (Mdn = 44, M = 44.1, and SD = 9.9). (Note the similarity between this distribution and the distribution in Study 1.) The BFS exhibited good internal reliability (Cronbach's α = .86). The test-retest correlation of total BFS scores was very strong (r = .88). This value exceeds the recommended benchmark of .70 for short-term test-retest reliabilities (Joiner, Walker, Pettit, Perez, & Cukrowicz, 2005).

We used a principal components analysis to identify potentially distinct subscales among the 18 BFS items. Varimax (orthogonal) and Promax (oblique) rotations produced nearly identical solutions. Both solutions revealed four factors with eigenvalues greater than 1.00, accounting for 58.3% of the total variance.

The first factor, which accounted for 33% of the variance, was identifiable as "belief in a controlling higher power." Items that loaded heavily on this factor—7, 9, 13, 15, and 17—mention God, a higher power, or a master plan. This 5-item subscale demonstrated very good internal consistency (α = .90). As expected, scores on this subscale correlated strongly (r = .82) with religiosity scores.

The second factor, which accounted for 11% of the variance, was identifiable as "belief in inevitability." Items that loaded heavily on this factor—1, 5, 8, and 12—reflect a *que sera sera* (whatever will be will be) view of the world. This 4-item subscale demonstrated acceptable internal consistency ($\alpha = .76$).

The third factor, which accounted for 7.5% of the variance, appears to capture feelings of resignation. Items that loaded heavily on this factor—10, 11, and 14—reflect concerns that "what I do does not matter." This 3-item subscale demonstrated mediocre internal consistency ($\alpha = .58$).

The fourth factor, which accounted for 7% of the variance, is less easily interpreted but appears to capture beliefs about locus of control. Items that loaded heavily on this factor—2, 4, 6, and 18—reflect beliefs about personal causality and external causality. This 4-item subscale demonstrated poor internal consistency ($\alpha = .39$). Our judgment that this subscale taps LOC is supported by the fact that its scores correlate more strongly with I-E scores ($r = .43$) than with BFS scores ($r = .25$).

BFS scores were strongly correlated with religiosity, $r(331) = .67$, $p < .001$. This raises the possibility that both scales actually measure the same thing. If so, we would expect the measures to operate similarly with respect to other variables, but that was not the case. BFS scores were positively correlated ($r = .21$, $p < .001$) with I-E scores, but religiosity scores were *negatively* correlated with I-E scores ($r = -.16$, $p < .01$). Moreover, BFS scores predicted binge-and-purge eating and social class (rs = .15 and .12, $p < .05$), but religiosity scores did not (rs = -.01 and -.05, n.s.). In sum, the constructs of fatalism and religiosity appear to overlap somewhat, but the BFS is not a measure of religiosity.

Women, as a group, are more fatalistic than men (Caplan & Schooler, 2003; Goodwin & Allen, 2000; Grassi et al., 2005; Roberts et al., 2000). Therefore, on the BFS, we expected women to score higher than men—and they did (Ms = 45.8 and 40.9, respectively), $t(328) = 4.39$, $p < .001$. Older people, as a group, are more fatalistic than younger people (Caplan & Schooler, 2003; Goodwin et al., 2002; Grassi et al., 2005). Therefore, we expected BFS scores to be positively correlated with age—and they were ($r = .11$, $p < .05$). I-E scores, however, were *negatively* correlated with age ($r = -.22$, $p < .01$).

As noted earlier, fatalistic thinking can be a *kind* of external LOC but is not synonymous with external LOC. As expected, BFS and I-E scores were correlated ($r = .21$, $p < .001$), but the modesty of the relationship indicates the two scales measure different constructs.

This inference was further supported by an examination of the relationships between BFS scores, I-E scores, and political party affiliation. If fatalism is synonymous with external LOC, then a political group that is high in fatalism should also be high in external LOC, but that was not the case. As shown in Table 14.1, Republicans scored highest on fatalism but lowest on external LOC. Greens scored lowest on fatalism but highest on external LOC.

Table 14.1 Means (and standard deviations) of BFS and I-E scores by political party affiliation.

Political party	Beloit Fatalism Scale	Rotter's I-E Scale
Republican (n = 37)	**50.4** [a] (10.1)	9.5 [a] (3.8)
Democrat (n = 208)	43.1 [b] (9.3)	11.1 [b] (4.1)
Green (n = 10)	41.2 [b] (8.6)	**13.5** [c] (3.7)
Independent (n = 43)	44.1 [b] (10.1)	10.6 [a,b] (3.9)
Other (n = 34)	44.8 [b] (11.6)	12.7 [c] (4.3)

Note. BFS = Beloit Fatalism Scale; I-E Scale = Rotter's Internal-External Locus of Control Scale. Highest mean score in each column is in boldface. Means with different superscripts in the same column are significantly different at $p < .05$ using Tukey's LSD test.

Two studies have found an inverse relationship between external LOC and subjective happiness (Lu, 1999; Pannells & Claxton, 2008). As expected, participants' I-E scores were negatively correlated with their Subjective Happiness scores (r = -.19, $p < .001$). Their BFS scores, however, were not correlated with happiness scores (r = .07, n.s.), which is additional evidence that the BFS and I-E Scale measure different constructs.

As shown in Table 14.2, BFS scores—but *not* I-E scores—predicted the self-reported frequency of skydiving, binge-and-purge eating, eating fast foods, and using seat belts.

Table 14.2 Spearman's rank correlation coefficients (*rho*) of BFS scores and I-E scores with self-reported protective behaviors.

Self-reported Behavior	Beloit Fatalism Scale	Rotter's I-E Scale
Frequency of Binge-and-Purge Eating	.15**	.07
Frequency of Eating Fast Foods	.15**	.02
Frequency of Looking Both Ways before Crossing Street	-.08	.10
Frequency of Skydiving	-.18**	-.03
Frequency of Flossing Teeth	.07	-.12*
Frequency of Using Seat Belts	.15**	-.09

Note. BFS = Beloit Fatalism Scale; I-E Scale = Rotter's Internal-External Locus of Control Scale.

**$p < .01$, *$p < .05$

General Discussion

The Beloit Fatalism Scale appears to be a useful measure of fatalism. The scale demonstrates good internal consistency, excellent test-retest reliability, and good discriminant validity. It produces a wide range of scores, is available at no cost, and is easy to administer.

Our findings strongly suggest that fatalism and external LOC are separate constructs. BFS and I-E scores do not correlate strongly with each other, and their correlations with other variables—age, happiness, political party affiliation, and self-reported protective behaviors—show different patterns. Given that fatalistic thinking is only one facet of an external LOC, researchers may wish to measure fatalism more directly with the BFS.

Our findings also raise basic questions about the underlying psychological structure of fatalism. In our sample, fatalism was comprised of four specific components: (1) belief in a controlling higher power, (2) belief in the inevitability of events, (3) a sense of resignation, and (4) an external locus of control. The presence of the first factor explains the strong association between fatalism and religiosity in our sample; in fatalistic thinking, there is a distinct element of *something or someone else* (usually a higher being) that controls events. Nevertheless, it is possible to be fatalistic without being religious. In our sample, participants' religiosity scores and scores on the BFS's second factor (inevitability) were correlated only modestly (r = .33), which means a good number of participants were fatalistic (that is, believed in the inevitability of events) without being especially religious (and vice versa). The presence of the third factor (resignation) is consistent with the oft-posited relationship between fatalism and depression (Neff & Hoppe, 1993; Roberts et al., 2000). A sense of resignation or helplessness may be the "active ingredient" in fatalism that leads to depressive symptoms. Finally, the presence of the fourth factor (external LOC) supports the claim made by many that fatalism and external LOC are related. The relationship, however, is not—and logically should not be—strong, given that fate is just one of several external causal factors.

The BFS demonstrated a modest degree of predictive validity. Fatalists, when compared to non-fatalists, were just slightly more likely to binge-and-purge or eat fast food on a regular basis. This did not surprise us for two reasons. First, almost all participants reported that they regularly engaged in healthy, protective behaviors, thus producing a restricted range problem. Second, if our conceptualization of fatalism as a general belief is valid, the predictive validity of the BFS should be modest because (a) most actions are multiply-determined and (b) daily decisions and general beliefs operate at maximally different levels of specificity (Fishbein & Ajzen, 1975).

Limitations and Directions for Further Research

The method used in the second study—an on-line survey with snowball sampling—produced at least two limitations. First, we were unable to observe and measure actual behavior. Second, our validation sample was limited in several respects. Ethnic minorities and men were underrepresented, while highly educated persons were overrepresented. Thus, it will be important to further validate the BFS using larger, more representative samples and different national samples. Indeed, cross-cultural studies may benefit from a standard measure that can be easily translated for use in different nations. With a common metric, researchers can observe regularities and patterns in fatalistic thinking and, as a result, gain a fuller understanding of fatalism as a psychological construct.

Finally, we wish to caution against the careless use of composite scores. Given our assumption that fatalism is a multi-faceted construct, we expected the BFS to be factorially complex—and it is. The disadvantage of a heterogeneous measure, of course, is that the same total score can be achieved in different ways, as is the case with other broad measures of multi-faceted constructs (intelligence, for example). Thus, we recommend the use of separate facet scores in addition to total BFS scores.

To illustrate this point empirically, we administered the BFS to American and Moroccan university students in a 2010 pilot study. According to many commentators, Muslim Arabs (including Moroccans) are more fatalistic than non-Muslims and non-Arabs (De Atkine, 2004; Elder, 1966; Nydell, 2005). The Moroccans in our study scored significantly higher than Americans on the first three facets, but they scored lower on the fourth facet—external locus of control.[3] Moroccans were more likely than Americans to agree with statements such as "people's misfortunes result from the mistakes they make" and "my future depends on the plans I make for myself." The Moroccan variety of fatalism appears to differ from the American variety, but this was only revealed by examining lower-level facet scores.

Our informants told us that, in Morocco, one can be fatalistic *and* possess an internal locus of control, as taught in a well-known proverb: A man came to visit the Prophet Mohamed and neglected to tie up his camel. When the Prophet asked him why he did not tie his camel, the man said, "There is no need. The Koran says to depend on God." The Prophet frowned and said, "First tie it up, then depend on God."

References

Caplan, L. J., & Schooler, C. (2003). The roles of fatalism, self-confidence, and intellectual resources in the disablement process in older adults. *Psychology and Aging*, *18*, 551–561.

Boone, C., de Brabandeer, B., Gerits, P., & Willeme, P. (1990). Relation of scores on Rotter's I-E scale to short-term and long-term control expectancies and fatalism. *Psychological Reports*, *66*, 1107–1111.

[3] In the pilot study, Moroccans were substantially more fatalistic (M = 60.5) than Americans (M = 46.1) overall, $t(243) = 8.17$, $p < .001$.

Cohen, D., & Nisbett, R. E. (1998). Are there differences in fatalism between rural Southerners and Midwesterners? *Journal of Applied Social Psychology, 28,* 2181–2195.

Cole, D., Rodriguez, J., & Cole, S. (1978). Locus of control in Mexicans and Chicanos: The case of the missing fatalist. *Journal of Consulting and Clinical Psychology*, *46*, 1323–1329.

Dake, K. (1992). Myths of nature: Culture and the social construction of risk. *Journal of Social Issues, 48*, 21–37.

De Atkine, N. B. (2004). The Arab mind revisited. *Middle East Quarterly*, *11*, 47–55.

Elder, J. W. (1966). Fatalism in India: A comparison between Hindus and Muslims. *Anthropological Quarterly*, *39*, 227–243.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.

Goodwin, R., & Allen, P. (2000). Democracy and fatalism in the former Soviet Union. *Journal of Applied Social Psychology, 30,* 2558–2574.

Goodwin, R., Allen, P., Nizharadze, G., Emelyanova, T., Dedkova, N., Saenko, Y., et al. (2002). Fatalism, social support, and mental health in four former Soviet cultures. *Personality and Social Psychology Bulletin, 28,* 1166–1171.

Grassi, L., Buda, P., Cavana, L., Annunziata, M. A., Torta, R., & Varetto, A. (2005). Styles of coping with cancer: The Italian version of the Mini-Mental Adjustment to Cancer (Mini-MAC) Scale. *Psycho-Oncology*, *14*, 115–124.

Joiner, T. E., Jr., Walker, R. L., Pettit, J. W., Perez, M., & Cukrowicz, K. C. (2005). Evidence-based assessment of depression in adults. *Psychological Assessment*, *17*, 267–277.

Koller, M. (2006, May). *Does the Future factor of the Zimbardo Time Perspective Inventory measure orientation toward the future or conscientiousness?* Paper presented at the 78th Annual Convention meeting of the Midwestern Psychological Association, Chicago, Illinois.

Lu, L. (1999). Personal or environmental causes of happiness: A longitudinal study. *Journal of Social Psychology*, *139*, 79–90.

Lyubomirsky, S., & Lepper, H. S. (1999). A measure of subjective happiness: Preliminary reliability and construct validation. *Social Indicators Research*, *46*, 137–155.

Matte, D., & Schaefer, C. S. (2004). An investigation of validity of the Subjective Happiness Scale. *Psychological Reports*, *94*, 288–290.

McClure, J., Allen, M. W., & Walkey, F. (2001). Countering fatalism: Causal information in news reports affects judgments about earthquake damage. *Basic and Applied Social Psychology, 23,* 109–121.

Neff, J. A., & Hoppe, S. K. (1993). Race/ethnicity, acculturation, and psychological distress: Fatalism and religiosity as cultural resources. *Journal of Community Psychology*, *21*, 3–20.

Nydell, M. K. (2005). *Understanding Arabs: A guide to modern times*. Boston: Intercultural Press.

Pannells, T. C., & Claxton, A. F. (2008). Happiness, creative ideation, and locus of control. *Creativity Research Journal*, *20*, 67–71.

Pearlin, L. E., & Schooler, C. (1978). The structure of coping. *Journal of Health and Social Behavior, 19,* 2–21.

Powe, B. D., Daniels, E. C., & Finnie, R. (2005). Comparing perceptions of cancer fatalism among African American patients and their providers. *Journal of the American Academy of Nurse Practitioners*, *17*(8), 318–324.

Realo, A., Allik, J., & Greenfield, B. (2008). Radius of trust: Social capital in relation to familism and institutional collectivism. *Journal of Cross-Cultural Psychology*, *39*, 447–462.

Roberts, R. E., Chen, C. R., & Ger, I. (2000). Fatalism and risk of adolescent depression. *Psychiatry, 63,* 239–252.

Ross, C. E., Mirowsky, J., & Cockerham, W. C. (1983). Social class, Mexican culture, and fatalism: Their effects on psychological distress. *American Journal of Community Psychology*, *11*, 383–399.
Rotter, J. (1966). Generalized expectancies for internal versus external control of reinforcements. *Psychological Monographs*, *80*, Whole No. 609.
Wade, T. J. (1996). An examination of locus of control/fatalism for blacks, whites, boys, and girls over a two-year period of adolescence. *Social Behavior and Personality, 24,* 239–248.
White, L. T., Valk, R., & Dialmy, A. (2011). What is the meaning of "on time"? The sociocultural nature of punctuality. *Journal of Cross-Cultural Psychology*, *42*, 482–493.
Zimbardo, P. G., & Boyd, J. N. (1999). Putting time in perspective: A valid, reliable individual-differences metric. *Journal of Personality and Social Psychology*, *77*, 1271–1288.

Appendix 14A Beloit Fatalism Scale.

No	Item
1	I believe in fate and destiny.
2	People's misfortunes result from the mistakes they make. (R)
3	Everything happens for a reason.
4	My future depends on the plans I will make for myself. (R)
5	I have often found that what is going to happen will happen.
6	Many times I feel that I have little influence over the things that happen to me because there are powerful forces at play.
7	I think that God or a higher power controls my life.
8	I think that if something is meant to be, it will be.
9	I believe that my life is part of a greater master plan
10	It does not make sense to worry about the future, since there is nothing I can do about it anyway.
11	It does not really matter what I do since whatever will be will be.
12	Fate determines much in my life.
13	To avoid misfortune, the will of God or a higher power is important.
14	The future is too unknown for a person to make serious plans.
15	Luck is really God or a higher power making something good happen.
16	My life path has been predetermined.
17	God works in mysterious ways.
18	Nothing is inevitable. (R)

Note. Agreement with each item is indicated on a 5-point Likert scale, 1 = *strongly disagree* and 5 = *strongly agree*. Items 2, 4, and 18 are reverse-scored (R). Items 2 and 5 come from Rotter's (1966) I-E Scale. Items 10, 11, and 12 are based on items from the Present-Fatalistic subscale of the Zimbardo Time Perspective Inventory (Zimbardo & Boyd, 1999). Item 14 is based on an item from Dake's (1992) scale. Remaining items were constructed by the authors.

15 ALLIK—A COMMON SURNAME IN SOUTHERN ESTONIA?

MARKKU VERKASALO

I studied the most frequent surnames in Southern Estonia using different methods. In 2004 and 2005, I tracked the frequency of surnames in the Southern Estonian phonebook. In 2005, I asked 14 students of psychology about the most common surnames using a free-format list and a given list of names. In 2018, I accessed the Statistics Estonia database to find the most common surnames in Estonia overall and Southern Estonia in particular. The results showed that the 2004 Southern Estonian phonebook enabled very good estimations of the most popular surnames in Southern Estonia. The students correctly identified five of the most common names using the free-format list. Those names were the same as for Estonia as a whole. The students overestimated the popularity of the name Allik—they ranked Allik as the fifth most popular surname in Southern Estonia; however, the real rank is about 60th.

Introduction

In 2004 and 2005, I was giving lectures to psychology students at the University of Tartu as part of the Erasmus Programme. One evening during my visit in 2004, I was in my hotel room with nothing else to read than a book of Estonian-Lithuanian-Latvian poetry and a Southern Estonian phonebook: *Lõuna-Eesti Telefonikataloog 2004*. As I am interested in statistics and research methods, I set a challenge for myself: could I obtain a rough understanding of the most common surnames in Southern Estonia in one hour? At this time open data via the internet were not as common as they are nowadays. Thus, I used the open data that I had in the phonebook.

I went through the phonebook in one hour using a system that I describe in the methods section. In the spring of 2005, I returned to teach in Tartu. I then expanded my research to encompass the level of knowledge that Estonian students have of the most common surnames in Southern Estonia. I asked 14 students of psychology their opinion on the most common surnames in two ways: with a free-format list and with a given list. That same week as I was in Tartu, an Estonian evening newspaper published an article about the 25 most common surnames in Estonia. Using all these data, I

I am grateful for my friendship with Anu Realo, Jüri Allik, and their daughter Annamari. I would like to thank them for all the hospitality they have shown me during my visits to Estonia and for their excellent lessons at my courses at the University of Helsinki. I would also like to thank Pekka Lahti-Nuuttila for his comments and Julie Uusinarkaus for her revision of this chapter.

E-mail: markku.verkasalo@helsinki.fi

wrote an entrance test for applicants wishing to study psychology at the University of Helsinki in the spring of 2005.

After I was invited to write this short chapter for the Festschrift in honor of Jüri Allik, in October 2018, I expanded my study with new data from the official database of Statistics Estonia[1]. I was now able to obtain the official statistics on all surnames in Estonia, updated as of 1 January 2018. I could also select statistics on different surnames in Southern Estonia from this data. I used the same definition for Southern Estonia as used in the phonebooks published in 2004 and 2005.

Southern Estonia consists of the provinces Jõgeva, Põlva, Tartu, Valga, and Võru. Statistics Estonia did not have data readily available for Southern Estonia as defined in the phonebook *per se*: I had to calculate the rates of the names in the different provinces and sum them up myself.

With all this information, I was able to formulate my research questions, as follows:

(a) How valid was the information obtained in one hour from the phonebook published in 2004 on Southern Estonian surnames compared with the official statistics from Statistics Estonia in 2018?
(b) How well do Estonian students know the most common surnames in Estonia, and can they identify these names, especially in Southern Estonia?
(c) Did I obtain more information from the phonebook in one hour than what Estonian students already knew about the most common surnames?
(d) Regarding the inquiry directed to the students: does a given list yield more exact results than a free-format list? As I also included the surname Allik in the given list, does a familiar name lead to any overestimation effect?
(e) What names are popular in Estonia as a whole and which are especially popular in Southern Estonia? Further, I became interested in the question: what categories of names are given as surnames in Estonia?

Methods and Data

Dataset A: The most common surnames in the Southern Estonian Phonebook 2004 [Lõuna-Eesti Telefonikataloog 2004]. From the Southern Estonia 2004 phonebook I selected the section of private phone numbers. These were all fixed line phone numbers; mobile phone numbers were not included in this phonebook. Each page had four columns, and each column had about 100 phone numbers. Next, I divided a page horizontally into 10 sections. I made a list of the most common surnames, that is, those which were listed at least 100 times in the phonebook. I rounded the name counts to the nearest 10. For example, the most common last name, Tamm, filled 4 columns and 2/10th of the fifth column on a page, so the estimated number was 420 surnames. After going through the whole phonebook, I found 31 surnames that had

[1] www.stat.ee

at least 100 listings. (Later in the chapter, I refer to the *Southern Estonia Phonebook 2004* as SE_PBK04).

Dataset B: The most common surnames in the Southern Estonian Phonebook 2005 [Lõuna-Eesti Telefonikataloog 2005]. When I visited Tartu in 2005, I was given the 2005 Southern Estonian phonebook as a present from the hotel. To check the reliability of my 2004 phonebook data, I calculated new rates for each last name that I had obtained from the 2004 phonebook. (Later in the chapter I refer to the *Southern Estonia Phonebook 2005* as SE_PBK05).

Dataset C: Students' evaluations of the most common surnames in Southern Estonia in the spring of 2005. I questioned 14 students of psychology about the most common last names in Southern Estonia in two ways:

C1: Free-format list. I gave students a paper with ten lines but no names. Students had to specify what they thought were the top 10 most common surnames in Southern Estonia and write them down in order from the most common to the 10th most common. I then scored the names, giving a name with rank one ten points, a name with rank two nine points, and so on, until the name that was ranked 10th received one point. (Later in the chapter, I refer to the free-format list as SE_STUFree05).

C2: Given list. After the students had written down their estimation of the 10 most common surnames, they turned over the paper. On the other side was a list of 30 names and a column with no numbers for each name. (Later in the chapter, I refer to the given list as SE_STUGiven05). For the list, I selected names from three categories from the *Southern Estonia Phonebook 2004* (SE_PBK04):

C2.1: The 10 most common surnames. These names had from 160 to 420 phone numbers.

C2.2: Ten quite common surnames. These names had from 60 to 160 phone numbers.

C2.3: Ten quite rare surnames. These names had from 10 to 40 phone numbers.

I also included in this third list the names of some lecturers in the Department, such as Allik (40 phone numbers) and Realo (5 phone numbers). The names in the list were in alphabetical order. The following are the surnames of lecturers included, with their respective category in brackets (the categories were not on the original list): Allik (3), Ilves (1), Ivanov/a (2), Juhkam (2), Kaasik (2), Kask (1), Kroon (3), Kukk (1), Kuusik (2), Kuznetsov (2), Kõiv (1), Leis (3), Lill (2), Luik (1), Puusepp (1), Pärn (2), Raudsepp (1), Realo (3), Rebane (1), Roos (2), Rõõmus (3), Saar (1), Sauk (3), Sikk (2), Suits (3), Tamm (1), Tammeorg (3), Uibo (2), Valk (3), and Viin (3). Students had to combine the following numbers with the above given list of names; the numbers were told to represent the number of phone numbers for a given name: 420, 280, 220, 200, 200, 200, 190, 170, 170, 160, 140, 140, 120, 120, 100, 100, 80, 80, 60, 60, 40, 40, 30, 30, 30, 20, 20, 10, 10, and 5.

Dataset D: The official figures on 1 January 2018 from Statistics Estonia. Statistics Estonia (www.stat.ee) has lists available of the 50 most common surnames for males

and females. I identified these 50 names from the database. Furthermore, the databases of Statistics Estonia also enabled me to look up any surname and view the rates for that name. I sought figures for women and men in Estonia as a whole for all the surnames included in the above datasets. (Later in the chapter, I refer to this dataset as EE_STAT18).

The Statistics Estonia platform also had a graph of the frequency of names per 10,000 inhabitants in all counties of Estonia, including Southern Estonian counties. Thus, I sought the frequencies for Southern Estonian counties and made a sum variable representing the number of Southern Estonian surnames. (Later in the chapter, I refer to this dataset as SE_STAT18).

Results

Figure 15.1 shows the 50 most common surnames in Southern Estonia in 2018, and the rank of these names in the entire country. The most important difference between all Estonian names and Southern Estonian names is that Russian names have much lower rates in Southern Estonia than in Estonia as a whole. Ivanov/a was the only Russian last name among the most popular Southern Estonian surnames, although in the list of all Estonians, the most common names include 13 Russian names. From the plot in Figure 15.1, we can see that the names that are clearly more popular in Southern Estonia than Estonia as a whole are Ilves, Rebane, Kukk, and Raudsepp, as well as, especially, Oja, Pärn, Kõiv, Põder, Lõhmus, Teder, and Kuusik.

The first research question was related to the validity of the information I obtained on Southern Estonian surnames from the *Southern Estonia Phonebook 2004*. As shown in Table 15.1, the Pearson correlation between the surnames in the Southern Estonian 2004 and 2005 phonebooks was very high, $r = .97$. The correlations between the results from Statistics Estonia in 2018 on Southern Estonian surnames (SE_STAT18) and all Estonian surnames (EE_STAT18) with the Southern Estonian phonebook in 2004 were $r = .92$ and $r = .97$, respectively. These highly significant correlations show that the *Southern Estonian Phonebook 2004* gave very good information about the most popular names in Southern Estonia, and that these results were very reliable.

When I studied the common name knowledge of Estonian students, I found that the correlations with the official Estonian (EE_STAT18) and the Southern Estonian (SE_STAT18) rates were also very good, with correlations being as high as rs= .78 and .70 with the free-format list, and .87 and .80 with the given list, respectively. The students' estimates of the frequency of Southern Estonian family names were also correlated with the frequency list obtained from the *Southern Estonia Phonebook 2004*, rs = .67 and .77 for the free-format and given lists, respectively. It is notable that students ranked Allik as 5th in the free-format list, even though its real rank is much lower (less than 60th), and as 11th in the given list, although its real rank was 20th in the 30 names given.

Table 15.1 Pearson correlations of ranks or frequencies of the surnames in Southern Estonia and throughout Estonia by different methods.

	EE_ STAT18	SE_ STAT18	SE_ PBK04	SE_ PBK05	SE_ STUFree05
EE_STAT18					
SE_STAT18	.96				
SE_PBK04	.92	.97			
SE_PBK05	.94	.99	.97		
SE_STUFree05	.78	.70	.67	.71	
SE_STUGiven05	.87	.80	.77	.79	.82

Note. $N = 28$. All correlations higher than $r = .55$ are significant at $p < .001$. EE_STAT18 = Estonia 2018, Statistics Estonia; SE_STAT18 = Southern Estonia 2018, Statistics Estonia; SE_PBK04 = *Southern Estonia Phonebook 2004*; SE_PBK05 = *Southern Estonia Phonebook 2005*; SE_STUFree05 = Southern Estonia free-format list 2005; SE_STUGiven05 = Southern Estonia given list 2005. In EE_STAT18, only Estonian names are included. If we include Russian names, the correlations of Estonian names with other variables are lower, with a range of .13 to .19.

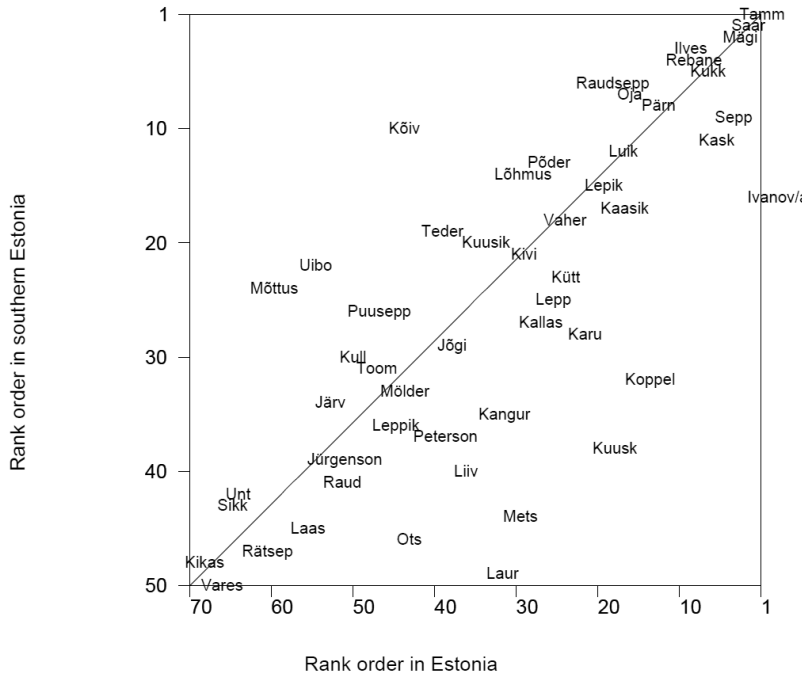


Figure 15.1 The 50 most common surnames in Southern Estonia (y-axis) and their rank in Estonia as a whole (x-axis) in 2018, according to Statistics Estonia.

Next, I compared the 12 most popular surnames in Southern Estonia according to Statistics Estonia in 2018 (SE_STAT18) with the student's free-format list from 2005 (SE_STUFree05) in order to examine the overlap between the two lists. The students' free-format list in 2005 had five names correct – Tamm, Saar, Mägi, Sepp, and Kask – which are indeed among the most popular family names in Estonia. In contrast, the students did not rate as high those names which belong to the top eight in Southern Estonia but which are not quite as popular in all of Estonia, such as Ilves, Rebane, Kukk, and Raudsepp.

When I compared the 12 most popular surnames in Southern Estonia according to Statistics Estonia in 2018 (SE_STAT18) with the list of the 12 most frequent family names in the Southern Estonia phonebooks of 2004 and 2005, there was an overlap in eight names with the 2004 phonebook (SE_PBK04), and all but one were among the 15 most highly ranked according to the SE_PBK04. As for the *Southern Estonia Phonebook 2005*, 11 were the same as the 12 most popular surnames in Southern Estonia according to Statistics Estonia.

Discussion

The results showed that the Southern Estonian phonebooks in 2004 and 2005 gave very good information about the most common surnames in Southern Estonia – the most frequent family names in these two phonebooks were also the most frequent family names in Southern Estonia in 2018 according to Statistics Estonia. The students knew the names in the top five list of all Estonian names, although very few of them wrote the names of Ilves, Rebane, Kukk, and Raudsepp, which are in the top eight of the most frequent family names in Southern Estonia. The Southern Estonian population is quite big, about a fifth of the population of all of Estonia, so even if Russian names are included, the 50 most popular Southern Estonian names can be found in the list of the top 70 Estonian most popular names. The Statistics Estonia website provided statistics on different provinces, but Southern Estonia as such was not defined. Thus, this study gives new information about the most popular last names in Estonia, and especially in Southern Estonia.

I also identified what categories of names are popular in Estonia in general. Nature seems to be an important source for Estonian surnames. It is notable that the same trend can be found in Finland, and to some extent also in Sweden. Among the 50 most popular Southern Estonian names, there are many references to natural settings, such as Mägi [*hill*], Oja [*creek*], Järv [*lake*], and so on (16 names). Then there were names based on types of tree, such as Tamm [*oak*], Kask [*birch*], Lepp [*alder*], and so on (eight names). In addition, there were names of animals, such as Ilves [*lynx*], Rebane [*fox*], Kukk [*rooster*], and Põder [*elk/moose*] (eight names). The fourth category was professions, such as Raudsepp [*blacksmith*], Sepp [*smith*], Puusepp [*carpenter*], Mölder [*miller*], Kangur [*weaver*], and Rätsep [*tailor*] (seven names).