19(4), pp.457-590, 2021 **ISSN 1334-4684**

INTERDISCIPLINARY DESCRIPTION OF COMPLEX SYSTEMS

Scientific Journal

R. Van Der Merwe	457	On Paul Cilliers' Approach to Complexity: Post-structuralism versus Model Exclusivity
A. Oblak	470	Participatory Sense-making as Consensual Validation of Phenomenal Data
S. Banerjee	493	An Artificial Immune System Approach to Automated Program Verification: Towards a Theory of Undecidability in Biological Computing
S. Banerjee	502	Lymph Node Inspired Computing: Towards Immune System Inspired Human-Engineered Complex Systems
I. Lipovac and M. Bagić Babac	511	Content Analysis of Job Advertisements for Identifying Employability Skills
Z. Šergo, J. Gržinić and A.S. Ilak Peršurić	526	The Effect of Tourism Overnight Stays on Croatia's Extra Virgin Olive Oil Prices and Market Power: An Empirical Study
R. Gsenger and T. Strle	542	Trust, Automation Bias and Aversion: Algorithmic Decision-Making in the Context of Credit Scoring
I. Saidi, I. Bejaoui and M.G. Xibilia	561	A Novel Discrete Internal Model Control Method for Underactuated System
Ž. Mekiš Recek, Z. Rojs, L. Šinkovec, P. Štibelj, M. Vogrin, B. Zamrnik and A. Slana Ozimič	580	Which Chord Progressions Satisfy us the most? The Effect of Expectancy, Music Education, and Pitch Height

Scientific Journal
INTERDISCIPLINARY DESCRIPTION OF COMPLEX SYSTEMS

INDECS, volume 19, issue 4, pages 457-590, year 2021
Published 30th December 2021 in Zagreb, Croatia
Released online 30th December 2021

Office

Croatian Interdisciplinary Society
c/o Faculty of Mechanical Engineering & Naval Architecture
I. Lučića 1, HR – 10 000 Zagreb, Croatia
E-mails: editor@indecs.eu (for journal), ured@idd.hr (for publisher)

Published by *Croatian Interdisciplinary Society* (http://www.idd.hr) quarterly as printed (ISSN 1334-4684) and online (ISSN 1334-4676) edition. Printed by *Redak d.o.o.* (HR) in 50 pieces. Online edition, http://indecs.eu, contains freely available full texts of published articles.

Journal INDECS is financially supported by Croatian Ministry of Science and Education.

Content of the journal INDECS is included in the DOAJ, EBSCO, EconLit, ERIH PLUS, Ulrich's and Web of Science Core Collection.

INDECS publishes original, peer-reviewed, scientific contributions prepared as reviews, regular articles and conference papers, brief and preliminary reports and comments to published articles. Manuscripts are automatically processed with the system Comet, see details here: http://journal.sdewes.org/indecs.

The accessibility of all URLs in the texts was checked one weak before the publishing date.

TABLE OF CONTENTS

Josip Stepanić ii INDECS Award

REGULAR ARTICLES

Ragnar Van Der Merwe 457 On Paul Cilliers' Approach to Complexity: Post-structuralism versus Model Exclusivity

Aleš Oblak 470 Participatory Sense-making as Consensual Validation of Phenomenal Data

Soumya Banerjee 493 An Artificial Immune System Approach to Automated Program Verification: Towards a Theory of Undecidability in Biological Computing

Soumya Banerjee 502 Lymph Node Inspired Computing: Towards Immune System Inspired Human-Engineered Complex Systems

Ivona Lipovac and Marina Bagić Babac 511 Content Analysis of Job Advertisements for Identifying Employability Skills

Zdravko Šergo, Jasmina Gržinić and Anita Silvana Ilak Peršurić 526 The Effect of Tourism Overnight Stays on Croatia's Extra Virgin Olive Oil Prices and Market Power: An Empirical Study

Rita Gsenger and Toma Strle 542 Trust, Automation Bias and Aversion: Algorithmic Decision-Making in the Context of Credit Scoring

Imen Saidi, Islem Bejaoui and Maria Gabriella Xibilia 561 A Novel Discrete Internal Model Control Method for Underactuated System

Žiga Mekiš Recek, Zala Rojs, Laura Šinkovec, Petra Štibelj, Martin Vogrin, Brina Zamrnik and Anka Slana Ozimič 580 Which Chord Progressions Satisfy us the most? The Effect of Expectancy, Music Education, and Pitch Height

INDECS AWARD

Dear authors of articles published in Vol. 18 of the journal INDECS,

the contest for the INDECS award, INDECSA 2021, choosing of the best article published in INDECS during 2021, i.e. in Vol. 19, is opened.

The voters are you, the authors of articles published in INDECS Vol. 18, i.e. in 2020, and the members of the INDECS' Editorial Board. Each and every voter contributes with one vote.

Propositions for the INDECSA are available from the web site of INDECSA, http://indecs.eu/index.php?s=indecsa.

I would like to ask you to give your vote to the article which you consider to be the best among the articles published in the year 2021.

The votes will be collected till 28th February 2022 and the results will be presented in INDECS 20(1).

Cordially,

Zagreb, 20th December 2021 Josip Stepanić

Interdisciplinary Description of Complex Systems 19(4), 457-469, 2021

ON PAUL CILLIERS' APPROACH TO COMPLEXITY: POST-STRUCTURALISM VERSUS MODEL EXCLUSIVITY

Ragnar Van Der Merwe*

University of Johannesburg, Faculty of Humanities, Department of Philosophy
Johannesburg, South Africa

DOI: 10.7906/indecs.19.4.1
Regular article

Received: 7 October 2021.
Accepted: 8 December 2021.

ABSTRACT

Paul Cilliers has developed a novel post-structural approach to complexity that has influenced several writers contributing to the current complexity literature. Concomitantly however, Cilliers advocates for modelling complex systems using connectionist neural networks (rather than analytic, rule-based models). In this article, I argue that it is dilemmic to simultaneously hold these two positions. Cilliers' post-structural interpretation of complexity states that models of complex systems are always contextual and provisional; there is no exclusive model of complex systems. This sentiment however appears at odds with Cilliers' promotion of connectionist neural networks as the best way to model complex systems. The lesson is that those who currently follow Cilliers' post-structural approach to complexity cannot also develop a preferred model of complex systems, and those who currently advocate for some preferred model of complex systems cannot adopt the post-structural approach to complexity without giving up the purported objectivity and/or superiority of their preferred model.

KEY WORDS

Paul Cilliers, Jacques Derrida, complexity theory, post-structuralism, connectionism, neural networks

CLASSIFICATION

JEL: C51

*Corresponding author, η: ragnarvdm@gmail.com; -;
University of Johannesburg, Faculty of Humanities, Department of Philosophy, Kingsway Campus, Corner Kingsway and University Road, Auckland Park, Johannesburg, 2000, South Africa

INTRODUCTION

It is generally recognised that one cannot model a complex system[1] without losing certain features of that system. Those who adopt a Derridean post-structural approach to complexity are particularly concerned with this loss – this *excess* – of meaning and therefore knowledge [1]. Meaning in and knowledge of complex systems cannot be reduced to some simple algorithm or set of rules; complex systems are informationally incompressible [2; pp.9-10, 3, 4]. In other words, complex systems cannot be reduced to simple models (otherwise they were not complex to begin with). Pockets of stability "make it possible to provisionally model a system", but "any model is contingent upon the context under which it is established" [3; p.9, 5].

It should follow from this line of thinking that post-structural complexity theorists do not attempt to develop a *specific* (i.e. non-provisional, non-contextual) model of complex systems. One must either argue for model provisionality, contingency, contextuality etc and forgo model exclusivity *or* advocate for some specific model and forgo model provisionality, contingency, contextuality etc. However, we find this ostensibly dilemmic approach in the work of Paul Cilliers who originated the post-structural approach to complexity. He argues both for post-structuralist provisionality, contingency, contextuality etc *and* that connectionist neural networks are the best way to model complex systems. Cilliers considers connectionist models superior to rival models of complex systems, and this is in tension with post-structural motifs of provisionality, contingency and the like, or so I will argue.

This dilemmic aspect of Cilliers' work has not been highlighted and thoroughly critiqued up until now, and this article should therefore make a novel contribution to the complexity literature. It should be of particular interest to those contemporary writers – e.g. Human [6], Hurst [7], Woermann [8] and Preiser [9] – who draw inspiration from Cilliers in continuing to develop the post-structural approach to complexity. It should also be of potential interest to those non-post-structural complexity theorists who currently advocate for some specific model of complex systems but may be considering adopting the post-structural interpretation.

The structure of this article is as follows. In section 1, I discuss how Derrida's semantics influences Cilliers' post-structural approach of complexity, specifically his modelling of complex systems. In section 2, I outline Cilliers' conclusion that connectionist neural networks better model complex systems than what he calls analytic or rule-based models. In section 3, I highlight the dilemma that follows from concurrently holding the views discussed in the previous two sections; I also respond to three potential counter-arguments. Lastly, I conclude by outlining what implications my argument has for those currently engaged in the debate over the modelling of complex systems. This conclusion is twofold:

1. Post-structural complexity theorists cannot propose that complex systems should be modelled some specific way rather than some other way.
2. Those who advocate for some specific model of complex systems – i.e. most scientists working on complex systems [10] – cannot adopt the post-structural interpretation of complexity without giving up the purported objectivity and/or superiority of their preferred model.

DERRIDA'S SEMANTICS AND CONNECTIONIST MODELLING

Cilliers notices a synonymy between Derridean post-structural semantics and connectionist neural network models. Both emphasise processes and relations; both are dynamic and qualitative. Conversely, what Cilliers calls the analytic or rule-based approach to modelling complex systems is static, reductionist, algorithmic and quantitative [2; Ch.1, 8; Ch.1]. Advocates of the analytic approach include Descartes, Newton, Chomsky, Fodor, Searle and

Habermas [2, 11]. For Cilliers, analytic approaches – reliant on strict measurement and deterministic rule-based methods – cannot model the subtle relational nature of truly complex systems such as language or the brain in the way that neural networks can.

Rule-based models are strictly formal; they conform to a precise logic and consist of sets of symbols standing in logical relations [2; Ch.1]. These symbols stand for only the 'important' parts of the system being represented, says Cilliers, resulting in invariable loss of fidelity. The behaviour of a complex system is simplified or reduced to a set of rules that attempt to describe the system. Rule-based models also require a central controller – the "meta-rules of the system" – that decides which rules should become active in the system. Importantly, "[i]f the central control fails, the whole system fails" [2; p.15].

Conversely, connectionist neural networks are modelled on the brain which consists of neurons and synapses in rich, informational interrelations. Neural networks contain multiple densely interconnected processing *nodes* (*viz.* neurons). Each node is influenced by and influences multiple other nodes. Nodes usually form three layers: the *input layer* that receives data to be processed by the network; the *output layer* that presents the output of the network's computations; and one or more *hidden layers* that form associations between the input layer and the output layer (and do not have any link to the outside of the network). Information flows from the input layer through the hidden layer/s to the output layer. According to Buckner and Garson,

> [i]f a neural net were to model the whole human nervous system, the input units would be analogous to the sensory neurons, the output units to the motor neurons, and the hidden units to all other neurons [12].

Each node (whether in the input, hidden or output layer) has a certain *activation value* determined by the information it receives. Above a certain threshold value, it will 'fire' and send information (determined by its input) to the next node; below the threshold value, it will remain dormant. The links between nodes have a certain numerical value or *weight* that represents the strength of that link. The sum of the inputs determines the output of the node which in turn influences the activation value of the next node and so on. All the nodes in the network are processing in parallel, and the values of the weights rather than features of the nodes determine the characteristics of the network.

When *training* neural networks, all the weights and thresholds are set to random values. Training examples are fed to the input layer and propagate through the network giving some random output. The weights and thresholds are then continuously adjusted until certain kinds of inputs reliably generate certain kinds of desired outputs. After some time, the network should be able to *generalize* these input/output computations to examples not in the original training set. Thus, concludes Cilliers,

> a network provided with enough examples of the problem it has to solve will generate the values of the weights by itself... It 'evolves' in the direction of a solution... The value of any specific weight has no significance; it is the patterns of weight values in the whole system that bear information. Since these patterns are complex, and are generated by the network itself... there is no abstract procedure available to describe the process used by the network to solve the problem. There are only complex patterns of relationships [2; p.28][2].

Let us now briefly survey Saussure's structural semantics (section 1.1), then look at Derrida's transformation of Saussure's structural semantics into a post-structural semantics (section 1.2). This exposition is necessary to understand which aspects of post-structuralism Cilliers

considers informative to complexity studies. Before turning to Cilliers' argument that connectionist models are superior to rule-based models of complex systems, we also discuss three core concepts Cilliers adopts from Derrida to inform his post-structural understanding of complex systems; these are openness, trace and *différance* (section 1.3).

SAUSSURE'S STRUCTURAL SEMANTICS

For Saussure [13] the meaning of a linguistic sign (composed of signifier and signified) is determined by how it *differs* from all the other signs in a linguistic system. We can think of a sign as a semantic node in a relational network. The sign does not determine the relations however; instead, the sign is the result of – it 'emerges' from – the relations. Further, the linguistic system changes as a result of its contingent and contextual use by a community of speakers and not by the decree of a central dictator or telos.

Saussure's influence has spread through the humanities [14]. Barthes [15] notably reinvented Saussurean signs as interwoven narratives or 'myths' that constitute the saturated cultural milieu surrounding us moment-to-moment. Saussure's linguistics, in its original form, has however fallen out of favour since the mid-20th century. The post-structuralist tradition in philosophy, of which Derrida and Cilliers are part, has – as the name suggests – largely superseded Saussure's structuralism[3].

For Cilliers, Saussurean models consisting of discrete signs are 'somewhat 'rigid' and

> Derrida's transformation of the system by means of a sophisticated description of how the relationships interact in time… provides us with an excellent way of conceptualising the dynamics of complex systems from a philosophical perspective [2, 16].

In Saussurean models each word has its place and its meaning in a mostly stable linguistic system. Although the system evolves, it remains in a relatively steady state near equilibrium. According to Cilliers, this is not how linguistic systems and complex systems in general behave. Derrida's critique and adaptation of Saussure better capture the non-linear and dynamic nature of complex systems [8; pp.134-135, 17; p.262].

DERRIDA'S POST-STRUCTURAL SEMANTICS

In Saussure's model the meaning of a sign is *present* to a speaker. The meaning of language is grounded in the subjectivity of the community of speakers using that language. Derrida argues however that the meaning of signs is ungrounded, unstable and unpredictable; there is always *excess of meaning*. As Cilliers puts it,

> the signified (or 'mental' component) never has any immediate self-present meaning. It is itself only a sign that derives its meaning from other signs. Such a viewpoint entails that the sign is, in a sense, stripped of its 'signified' component [2; p.42, 8; p.72].

For Derrida, there is only the endless interaction of signifiers (the 'physical' component of the sign), and the subject itself is constituted by this play of signifiers [2; p.43]. Meaning is never immediately given; there is always *interpretation*, and interpretation is always limited. This is Derrida's famous *deconstruction* of the sign [18][4].

Each time a sign is used, it interacts with the other nodes in the linguistic network, and this semantic interplay shifts the meaning of the sign [17]. For Derrida and Cilliers, language is in a sense *alive*. It mutates, adapts and evolves; it acts on and reacts to its environment (including other languages). Like any complex system, a living language is in a state far from

equilibrium, and if "language is closed off, if it is formalised into a stable system in which meaning is fixed, it will die..." [1, 2].

OPENNESS, TRACE AND DIFFÉRANCE

Openness

For Derrida and Cilliers, language and meaning are not closed off from the world; semantics cannot be pulled apart from metaphysics, and we cannot describe the world in any complete, finite way [2; Ch.3, 19; p.35]. The same applies to complex systems: we cannot identify their boundaries in a way that is objective or complete. Complex systems are entwined with their environment which is itself a complex system composed of complex systems.

Delineating complex systems involves only a provisional, conceptual or heuristic demarcation;

> [w]hat occurs inside our models cannot be easily separated from what is excluded because what we exclude from our models constitutes them as much as that which is included [6; p.9, 7, 10].

Citing Cilliers, Woermann et al state, "our models are distorted... models are static representations of a necessarily fluid reality" [3; p.10]. As mentioned in the introduction, for post-structural complexity theorists, we cannot get a semantic or epistemic fix on complex systems. Thus, instead of trying to decomplexify complexity, we should "abandon our reductionist tendencies" and "learn to dance with" complexity [5, 7, 8]. Post-structuralism suggests a "'playful' approach", writes Cilliers,

> [w]hen dealing with complex phenomena, no single method will yield the whole truth. Approaching a complex system playfully allows for different avenues of advance, different viewpoints, and, perhaps, a better understanding of its characteristics [2; p.23].

In other words, we cannot semantically or epistemically capture – i.e. model – complex systems in any general, perspective-independent way.

Trace

Derrida calls the relationship between any two signs in a semantic system a *trace*. An individual trace does not have meaning in and of itself; instead, meaning emerges through the interaction of traces [20; pp.3-27, 21]. Cilliers equates Derrida's traces with connectionist weights in a neural network:

> The significance of a node in a network is not a result of some characteristic of the node itself; it is a result of the pattern of weighted inputs and outputs that connects the node to other nodes [2; p.81].

Likewise, no individual weight in a neural network has meaning; meaning is constituted by multiple interactions in the system. "Because of the 'distributed' nature of these relationships, a specific weight has no ideational content"; it "only gains significance in large patterns of interaction" [2; p.46][5]. In other words, all the small meaningless differences between the many components in a complex system comingle to engender the emergence of meaning within the system.

According to Cilliers, the patterns of activity generated in a complex system cause traces of that activity to reverberate through the system. These patterns of traces collectively constitute

the overall behaviour of the system. Moreover, a complex system is continuously being transformed by both its environment and itself. The system is

> constituted only by the distributed interaction of traces in a network... there is nothing outside the system of signs which could determine the trace, since the 'outside' itself does not escape the logic of the trace [2; p.82].

This entwinement of system and environment *deconstructs* the conventional binary of inside versus outside the system; the traditional gap between the two collapses. That is, traces ripple and recoil – they dance – through the system; they are "reflected back after a certain propagation delay (deferral), and alter (make different) the activity that produced them in the first place" [2; p.46].

Différance

Although reluctant to define *différance*, Derrida suggests at times that his famous concept *qua* non-concept is "the process of scission and division... an expenditure without reserve, as the irreparable loss of presence... that apparently interrupts every economy" [20; pp.8-19], i.e. every complex system [6, 17]. For Cilliers, *différance* is

> a concept that indicates difference and deference, that is suspended between the passive and active modes, and that has both spatial and temporal components [20; pp.1-27, 21; p.7].

In the context of complexity theory, Woermann thinks of *différance* as the play of disorder and entropy within a complex system [8; p.64]. *Différance* constitutes the activity of multitudinous traces: the exuberant and limitless play of differences. *Différance* disrupts, displaces and defers apparent closure of order, logic, meaning and knowledge [3, 8; pp.100-104]. The play of *différance* through and between complex systems constitutes their meaning and this can never be epistemically captured by formal methods. *Différance* "signifies the irreparable loss of meaning"; it "threatens the total ruination of meaning" [8; p.100, 9].

For Cilliers, *différance* describes the *dynamics* of a complex system. It is not simply part of the activity of a system; "it constitutes the system" [21; p.15]. We can say that the play of *différance* determines the structure or organisation of the system; the complexity of the system is a function of *différance's* dynamics.

Having discussed how Cilliers imports Derrida's post-structural semantics into complexity theory, let us now look at his proceeding conclusion that connectionist models are better suited to modelling complex systems than rule-based models.

THEREFORE, CONNECTIONIST MODELS TRUMP RULE-BASED MODELS

Cilliers prefers (post-structural) connectionist models to (analytic) rule-based models because of their avowed ability to capture the contingent, evolutionary nature of complex systems. Moreover, neural networks are based on the most complex of all known systems: the brain [2; p.112][6]. Like complex systems, neural networks have no central controller; "[p]rocessing is distributed over the network and the roles of the various components (or groups of components) change dynamically" [2; p.19]. Neural networks can also *learn* to perform complex tasks either when shown examples of these tasks successfully performed, or by using criteria internal to the network that signal success[7].

Neural networks are mostly self-contained, says Cilliers, they require only a *sensor* that inputs information to the network and a *motor* that allows the output of the system to have

some external effect [2; p.18]. Inside the network there are only neurons responding to and influencing other neurons *locally*. The behaviour of the system is determined only by the values of its weights. Each neuron is simple, but the system of neurons as a whole can exhibit highly complex behaviour[8].

Neural networks can also cope with contradictory information; they are 'robust'. Part of the strength of neural networks, says Cilliers, is that they can often bypass a contradiction by redistributing the weight in the system [21; pp.19-21]. Rule-based systems conversely are "brittle"; they "blow up" when given contradictory information[9].

THE PROBLEM WITH CILLIERS' APPROACH TO COMPLEXITY

As we have seen, Cilliers argues that using neural networks is the best way to model complex systems while concurrently arguing that post-structuralist semantics shows that there are no general models for complex systems. In this section, I suggest that it is dilemmic to do so (section 3.1); I then engage with three potential counter-arguments (section 3.2).

CILLIERS' DILEMMA

As Cilliers recognises, modelling necessarily involves a simplification or reduction of the system being modelled; "we have to reduce ... complexity when we try to understand it" [6, 22]. *A fortiori*, this reduction applies equally to analytic *and* connectionist models. Although Cilliers does not explicitly state as much, connectionist modelling clearly involves a reduction of complex systems to simple or simpler neural networks (consisting of nodes, weights etc). As we have seen however, Cilliers advocates for this connectionist reductionism while concurrently advocating for Derridean anti-reductionism.

Cilliers states further that "complexity is 'incompressible' " [2; p.24]; "[r]eduction of complexity always leads to distortion" [23; pp.9-10]. However, we are also told that connectionist neural networks are the best way to model – i.e. compress/reduce – complex systems. Cilliers also argues that a post-structural understanding of complexity shows that reductive strategies are "seriously flawed" [8; pp.31, 9, 21]. If so, it follows that Cilliers' own connectionist strategy is seriously flawed, and thereby inept at modelling complex systems. De Villiers-Botha and Cilliers likewise argue that one cannot replace a complex system with some simpler system without losing "vital characteristics of the system" [9; p.29]. It should follow that connectionist models lose vital characteristics of their target system.

Cilliers claims further that models are always indexed to some contingent framework, and that the success of a model will depend on the norms or values operant in that framework [6, 24; pp.45-46]. There are no objectively correct models of complex systems; there are no meta-narratives expressed from meta-perspectives [25; pp.458-450]. However, at other times, Cilliers argues that we should employ

> post-structural perspectives (mainly those of Derrida) in order to show that the intricate and dynamic network of relationships between the components of a complex system can be understood better in terms of connectionist (or neural network) models [24; p.40].

Cilliers seem to be arguing both for *and* against contingency, perspectivism and objectivity about modelling complex systems[10]. This suggests a logical tension at the core of his general account of complex systems.

Interestingly, Holland argues that connectionist models are – at bottom – themselves rule-based; the functioning of neural networks is based on the workings of Hebb's rule [26; pp.81-114][11]. If so, then Cilliers' connectionism is subject the very criticism he levels against rule-based

models. Cilliers however argues against Holland that Hebb's 'rule' is, in fact, not a rule at all; it is instead a kind of low-level *principle*. Citing Winston [27] Cilliers posits two features that constitute a rule:

1. A rule implies a certain generality; a specific case where the rule applies must be generalisable to many cases.
2. Rules must be suitably linked: the output of one rule must serve as the input for the next rule. That is, a system of rules must be algorithmic [24; pp.44-45].

For Cilliers, this characterisation of a rule is at odds with a nonalgorithmic principle like Hebb's rule. Hebb's 'rule', says Cilliers, in fact functions at the level of Derrida's *trace*. It only applies to local interactions between components of a complex systems. It operates on low-level, contingent information, and is non-selective. The 'rule' operates everywhere in all connectionist networks; it does not tell us anything essential about a specific complex system nor make any generalisations about complex systems.

In any event, neural networks are still clearly simpler that the complex systems they purport to model; they are still a reduction of complexity [26; p.24]. If they were not, we would not be able to comprehend them or work with them; the model would be as complex as what it attempts to model. As noted, modelling – whether by way of analytic rules or neural networks – by definition involves a reduction of complexity so that the subject matter at hand can be understood and managed.

In sum, Cilliers faces a dilemma. On the one horn, if he claims that the connectionist approach to modelling complex systems is the correct one, then he contradicts post-structural themes of perspectivism, contingency and the like. On the other horn, if he claims that models of complex systems are relative to perspectives, always contingent and so on, then he contradicts his claim for the correctness of connectionist models of complex systems[12].

POSSIBLE RESPONSES

Cilliers may respond that connectionism is the best way to model complex systems given the scientific paradigm we currently occupy even if there is no absolute fact of the matter. There are no meta-models or meta-perspectives, but some norms or methods are more 'long-lived' than others due to provisional, contingent or heuristic factors [1, 7]. However, by *reductio*, it should then follow that our best scientific models – e.g. in quantum theory, general relativity and the theory of natural selection – are ultimately no better than those employed in pseudo-sciences such as astrology, Scientology or creation biology. All are in the end equally provisional; the only reason we prefer the former is due to our currently dominant contingent norms. It would be grossly counter-intuitive to assert as much; one wonders whether post-structural complexity theorists are prepared to bite this bullet.

A second possible objection is that there are two domains of applicability when it comes to understanding complex systems. At times, Cilliers follows Morin [10] in distinguishing between *restricted* complexity and *general* complexity[13]. Cilliers may state that connectionism applies to restricted complexity while post-structuralism applies to general complexity. These two domains involve adopting a certain approach or mode of thinking towards modelling complex systems. Regarding restricted complexity, says Cilliers, "if you work hard enough, with clever enough techniques, you can figure the system out" [21; p.7]. Conversely, general complexity "requires a more reflexive and transformative approach" where we remain sensitive to the recalcitrance of complex systems and the normativity this introduces [21; p.7]. As Woermann et al put it,

> [i]n the restricted paradigm, complexity is treated as a problem that can be overcome (complex problems are understood as complicated problems);

> whereas in the general paradigm, complexity is treated as an ontological fact, which holds certain epistemological and cognitive implications for the manner in which we deal with complexity [3; p.5].

In the restricted mode, complex systems are considered to be epistemically complex but ontologically simple; in the general mode, complex systems are considered to be both epistemically and ontologically complex. Restricted complexity applies to Saussure, Chomsky and other reductive strategists, while general complexity applies to Derrida and likeminded post-structuralists.

Thus, Cilliers may claim that connectionist models are useful in the restricted mode, while post-structural thinking is useful in the general mode, and my dilemma therefore does not bite. However, drawing such demarcations is at odds with post-structural themes of *différance* and deconstruction [7, 28]. As mentioned in the section 1.3, *différance* disrupts all (non-provisional/non-heuristic) distinctions. Post-structuralism disallows robust demarcations; all substantial binaries are vulnerable to deconstruction [8; pp.173-176, 28; p.116].

Thus, claiming that there are two separate domains – one applicable to connectionism and the other applicable to post-structuralism – violates post-structuralism's own taboo on strict demarcations. Cilliers cannot claim that connectionism only applies to restricted complexity. This is because the force of *différance* should disrupt any attempt at isolating or closing off some system or domain [3; pp.7-10, 6]. On the post-structuralists' own account, *différance* should render connectionist modelling efforts as contingent, incomplete and contextual as rule-based efforts. However, as discussed, Cilliers does not consider connectionism relativised in this way. Instead, he argues for the superiority of connectionist models over rule-based models *simpliciter*; the correctness of connectionism is a putative consequence of Derrida's semantics.

Moreover, as we saw in section 1, Cilliers' advocation of post-structuralism is premised in its *similarities* to connectionism and *vice versa.* Cilliers therefore cannot claim that post-structuralism and connectionism apply to separate domains since they would then be disanalogous. If post-structuralism and connectionism are distinct, then they cannot support each other in the kind of argument by analogy that Cilliers puts forward.

Lastly, Cilliers may respond that the dilemma at hand is exactly the sort of paradox post-structuralists revel in. According to Human and Cilliers, we deal with paradox by utilising a type of 'reason' "defined as a wager between the calculable and the incalculable" [6; p.34]; the post-structural approach "harbours a somewhat ironic dimension" [29]. This involves an *aporetic logic* – a kind of dialethism – premised on *différance* that embraces paradoxes and contradictions [8; pp.67-81, 7; pp.243-246, 28; p.116]. Here, we must concurrently think both closed *and* open, both random *and* predictable, p *and* ~p, true *and* false.

Perhaps appeal to *différance* with its aporetic logic dispels all putative dilemmas: precisely where there is contradiction there is epistemic illumination. As Woermann states, the relationship between the restricted and the general dimensions of complexity is "better understood in terms of *aporia* than unification" [8; p.73]. Further,

> [i]n trying to think together the restricted and general dimensions of meaning, Derrida's logic aims to transgress the limitations that our traditional binary logical schema (which is necessary restricted) places on us [8].

However, embracing contradiction to dissolve a dilemma potentially leads to a debilitating kind of relativism. It seems that no positive argument can sustain the force of *différance. Différance* ruins all non-aporetic logics – it disrupts meaning, order and structure [8, 20] and one therefore wonders what kind of statement can actually be meaningful or be known. If all positive claims are prone to disruption, then how can there be norms of epistemic

correctness? Any claim to knowledge introduces a binary between known and unknown that should *itself* be vulnerable to the disruptive power of *différance* [8; pp.174-175, 28; p.116][14]. The result seems to be a version of what Goldman [30] calls *nihilistic relativism*: there are no non-contextual epistemic norms governing which claims are right or wrong or more right or wrong than others.

CONCLUSION

The conclusion to this article is twofold:

1. Post-structural complexity theorists cannot advocate for a specific model of complex systems.
2. Conversely, complexity theorists who advocate for a specific model of complex systems cannot embrace the post-structural interpretation.

Regarding 1, to my knowledge, none of the complexity theorists who follow Cilliers' post-structural approach currently attempt to develop or advocate for a specific model of complex systems[15]. As such, my argument here amounts to a warning to these theorists that they should not attempt to develop any exclusive model of complex systems in the future. To do so would entail falling prey to Cilliers' dilemma.

Conclusion 2 however has more pertinent implications. Much of Cilliers' writings and the writings of those who follow his approach are taken up with arguing that complexity theorists who adopt the restricted mode of understanding complex systems should be cognisant of the general mode. In other words, complexity theorists working in the analytic, reductive traditional would do well to embrace a post-structural way of thinking that is sensitive to the provisionality, contextuality and contingency involved in modelling complex systems. However, if my argument above holds, such potential converts cannot do so without ceasing to hold to any specific model they may have already developed or subscribed to.

Most non-post-structural complexity theorists currently advocate for some preferred modelling method. The frontier of the discipline involves methodical testing and engaged debate over the best way to model complex systems. These modellers are often post-structural complexity theorists' target audience. One wonders however whether these modellers are aware that embracing the post-structural mode of thinking will necessitate giving up any claims to the objectivity or superiority of their favourite model.

In sum, those non-post-structural modellers who may find aspects of the post-structural approach convincing, and who may be considering adopting it, would do well to note the sacrifice their conversion would entail. Likewise, post-structural complexity theorists who aim to convince others to embrace the post-structural approach would do well to make explicit the sacrifice entailed therein.

REMARKS

[1]For the purposes of this article, we can follow Richardson and Cilliers in defining a complex system as "a system that is comprised of a large number of entities that display a high level of nonlinear interactivity" [31]. See however Cilliers and Preiser [2; Ch.1, 21] for detailed lists of characteristics of complex systems. We can further define 'complexity theory' broadly as any practice that involves the attempt to gain knowledge and/or understanding of complex systems. A 'complexity theorist' is someone who engages in such a practice.

[2]See Cilliers and Buckner and Garson for more detail on neural networks [2; Ch.2, 10].

[3]Lazard is however a notable exception. Using the Saussurean approach, he attempts to develop a science of linguistics or what he calls "pure linguistics" [32] where a language can be an ideal object of genuine scientific investigation, as are biological species, a pure body and a perfect gas.

[4]According to Cilliers, post-structural deconstruction involves showing the contradictions that result from fixing the boundaries [of a complex system] from one perspective. Pointing out the contradictions that follow from such a closure is an activity that Derrida calls 'deconstruction' [2; p.81]. Deconstruction is not an action performed by a deconstructor on a system. Instead, "[i]nterventions from the outside enter into the play of differences always already at work in the system" [21; p.15]. Derrida, says Cilliers, "sometimes refers to deconstruction as a characteristic of the system itself: *it* deconstructs" [21].

[5]These weights "store information at a sub-symbolic level, as *traces* of memory" [2].

[6]See Ladyman and Wiesner for an informative explication of the brain as a highly complex system [33; pp.57-61].

[7]Cilliers also emphasises two "indispensable capabilities" of complex systems. A complex system must "be able to store information concerning the environment for future use; and it must be able to adapt its structure when necessary" [2]. The first capability is *representation* and the second *self-organisation.* According to Cilliers, connectionist models of complex systems correctly capture these capabilities [2; Ch.1-2].

[8]To perform their function, neural networks also cannot be too small (too few neurons) or too large (too many neurons), and they must not be under- or over-trained. Some noise should also be added to the input of the network to increase its robustness. [2; pp.74-79].

[9]Smolensky [34] thinks of rule-based symbol systems as "hard" and connectionist systems as "soft" [2; p.34].

[10]Morçöl criticises Cilliers along similar lines [35; pp.117-118].

[11]Hebb's rule describes the local interaction between neurons responsible for the organization of structure in a neural network; it states that "the connection strength between two neurons will increase if the two neurons are active simultaneously" [23; p.44].

[12]Ladyman and Wiesner suggest that this kind of paradoxical thinking runs throughout the complexity literature [33; p.84].

[13]Morin's *general complexity/restricted complexity* distinction approximately maps onto the more familiar complex system/complicated system distinction [4, 36].

[14]As solution to this dilemma, Derrida appeals to a kind of mystical moral force in the world that serves as first philosophy [37]. Cilliers however, does not follow Derrida in this regard.

[15]Woermann does however devote one paragraph to what appears to be an endorsement of neural networks as an appropriate model for complex systems [8; p.29]. Nonetheless, other than this token gesture, she does not argue for connectionism at any length nor attempt to develop her own specific model for complex systems.

REFERENCES

[1] Cilliers, P.: *Knowledge, limits and boundaries.*
In: Preiser, R., ed.: *Critical Complexity: Collected Essays.* De Gruyter, Berlin, pp.105-114, 2016,

[2] Cilliers, P.: *Complexity and Postmodernism: Understanding Complex Systems.*
Routledge, London, 1998,

[3] Woermann, M.; Human, O. and Preiser, R.: *General Complexity: A Philosophical and Critical Perspective.*
Emergence: Complexity and Organization, 2018,
http://dx.doi.org/10.emerg/10.17357.c9734094d98458109d25b79d546318af,

[4] Preiser, R. and Woermann, M.: *Complexity, Philosophy and Ethics.*
In: Galaz, V., ed.: *Global Challenges, Governance, and Complexity: Applications and Frontiers.*
Edward Elgar Publishing, Northampton, pp.38-62, 2019,

[5] Cilliers, P.: *Complexity, Deconstruction and Relativism.*
Theory, Culture & Society **22**(5), 255-267, 2005,
http://dx.doi.org/10.1177/0263276405058052,

[6] Human, O. and Cilliers, P.: *Towards an Economy of Complexity: Derrida, Morin and Bataille.*
Theory, Culture and Society **30**(5), 24-44, 2013,
http://dx.doi.org/10.1177/0263276413484070,
[7] Hurst, A.: *Complexity and the Idea of Human Development.*
South African Journal of Philosophy **29**(3), 233-252, 2010,
http://dx.doi.org/10.4314/sajpem.v29i3.59144,
[8] Woermann, M.: *Bridging Complexity and Post-Structuralism: Insights and Implications.*
Springer, Cham, 2016,
[9] Preiser, R.: *Identifying General Trends and Patterns in Complex Systems Research: An Overview of Theoretical and Practical Implications.*
Systems Research and Behavioural Science **36**(5), 706-714, 2019,
http://dx.doi.org/10.1002/sres.2619,
[10] Morin, E.: *Restricted Complexity, General Complexity.*
In: Gershenson, C.; Aerts, D. and Edmonds, B., eds.: *Worldviews, Science and Us: Philosophy and Complexity.* World Scientific, Singapore, pp.5-29, 2007,
[11] Kauffman, S.: *A World Beyond Physics: The Emergence and Evolution of Life.*
Oxford University Press, New York, 2019,
[12] Buckner, C. and Garson, J.: *Connectionism.*
In: Zalta, E.N., ed.: *The Stanford Encyclopedia of Philosophy,* 2019,
http://dx.doi.org/10.4324/9781315643670,
[13] de Saussure, F.: *Course in General Linguistics.*
Fontana, London, 1974,
[14] Joseph, J.E.: *Trends in Twentieth-Century Linguistics: An Overview.*
In: Koerner, E.F.K. and Asher, R.E.: *Concise History of the Language Sciences: From the Sumerians to the Cognitivists.*
Pergamon Press, Oxford, 221-233, 1995,
[15] Barthes, R.: *Mythologies.*
J. Cape, London, 1972,
[16] Dillon, M.: *Poststructuralism, Complexity and Poetics.*
Theory, Culture and Society **17**(5), 1-26, 2000,
http://dx.doi.org/10.1177/02632760022051374,
[17] Preiser, R.; Cilliers, P. and Human, O.: *Deconstruction and Complexity: A Critical Economy.*
South African Journal of Philosophy **32**(3), 261-273, 2013,
http://dx.doi.org/10.1080/02580136.2013.837656,
[18] Lawlor, L.: *Jacques Derrida.*
The Stanford Encyclopedia of Philosophy, 2021,
[19] Derrida, J.: *Of Grammatology.*
Johns Hopkins University Press, Baltimore, 1976,
[20] Derrida, J.: *Différance.*
University of Chicago Press, Chicago, pp.1-28, 1982,
[21] Cilliers, P.: *Difference, Identity and Complexity.*
In: Cilliers, P. and Preiser, R., eds.: *Issues in Business Ethics: Complexity, Difference and Identity.*
Springer, London, pp.1-18, 2010,
[22] Cilliers, P. and Preiser, R.: *Preface.*
In: Cilliers, P. and Preiser, K.: *Issues in Business Ethics: Complexity, Difference and Identity.*
Springer, London, iv-ix, 2010,
[23] Cilliers, P.: *Knowledge, Complexity, and Understanding.*
Emergence **2**(4), 7-13, 2000,
http://dx.doi.org/10.1207/S15327000EM0204_03,

[24] Cilliers, P.: *Rules and Complex Systems.*
Emergence **2**(3), 40-50, 2000,
http://dx.doi.org/10.1207/S15327000EM0203_04,
[25] Woermann, M. and Cilliers, P.: *The Ethics of Complexity and the Complexity of Ethics.*
South African Journal of Philosophy **31**(2), 447-464, 2012,
http://dx.doi.org/10.1080/02580136.2012.10751787,
[26] Holland, J.H.: *Emergence: From Chaos to Order.*
Addison-Wesley, Boston, 1998,
[27] Winston, K.I.: *Justice and Rules: A Criticism.*
Logical Analysis **14**(1), 177-182, 1971,
[28] Derrida, J.: *Afterword.*
In: Graff, G., ed.: *Limited Inc.* Northern Western University Press, Evanston, pp.111-160, 1988,
[29] Human, O.: *Potential Novelty: Towards an Understanding of Novelty without an Event.*
Theory, Culture & Society **32**(4), 45-63, 2016,
[30] Goldman, A.: *Epistemic relativism and reasonable disagreement.*
In: Feldman, R. and Warfield, T.: *Disagreement.* Oxford University Press, Oxford, 187-215, 2010,
[31] Richardson, K. and Cilliers, P.: *Special Editors' Introduction: What Is Complexity Science? A View from Different Directions.*
Emergence **3**(1), 5-23, 2001,
http://dx.doi.org/10.1207/S15327000EM0301_02,
[32] Lazard, G.: *The Case for Pure Linguistics.*
Studies in Language **6**(2), 241-259, 2014,
[33] Ladyman, J. and Wiesner, K.: *What Is a Complex System?*
Yale University Press, London, 2020,
[34] Smolensky, P.: *The Constituent Structure of Connectionist Mental States: A Reply to Fodor and Pylyshyn.*
Southern Journal of Philosophy **26**, 137-161, 1987,
http://dx.doi.org/10.1007/978-94-011-3524-5_13,
[35] Morçöl, G.: *What Is Complexity Science? Postmodernist or Postpositivist?*
Emergence **3**(1), 104-119, 2001,
http://dx.doi.org/10.1207/S15327000EM0301_07,
[36] Poli, R.: *A Note on the Difference Between Complicated and Complex Social Systems.*
Cadmus **2**(1), 142-147, 2013,
[37] De Villiers-Botha, T. and Cilliers, P.: *The Complex 'I': The Formation of Identity in Complex Systems.*
In: Cilliers, P. and Preiser, R.: *Issues in Business Ethics: Complexity, Difference and Identity.* Springer, London, pp.19-38, 2010,
http://dx.doi.org/10.1007/978-90-481-9187-1_2.

Interdisciplinary Description of Complex Systems 19(4), 470-492, 2021

PARTICIPATORY SENSE-MAKING AS CONSENSUAL VALIDATION OF PHENOMENAL DATA

Aleš Oblak*

University Psychiatric Clinic Ljubljana, Laboratory for Cognitive Neuroscience and Psychopathology Ljubljana, Slovenia

DOI: 10.7906/indecs.19.4.2
Regular article

Received: 7 January 2020.
Accepted: 23 December 2021.

ABSTRACT

This article proposes a method for consensually validating phenomenal data. Such a method is necessary due to underreporting of explicit validation procedures in empirical phenomenological literature. The article argues that descriptive sciences – exemplified by phenomenology and natural history – rely on nominalization for construction of intersubjectively accessible knowledge. To this effect, epistemologies of phenomenology and natural history are compared. The two epistemological frameworks differ in terms of their attitudes towards the interpretation of texts and visual epistemology, however, they both rely on eidetic intuition of experts for knowledge construction. In developing the method of consensual validation, I started out with the prismatic approach, a method for researching embodied social dynamics. I then used debriefings on the experience of consensual validation to further refine the method. The article suggests that for a nominalization of experiential world to be intersubjectively accessible, such a vocabulary must be independently constructed by the entire group of co-researchers. I therefore propose that during consensual validation, co-researchers be presented with composite descriptions of experiential categories, compare them with their experience, attempt to falsify them, and finally jointly name them. This approach does not yield a single vocabulary for description of experience, but several commensurable vocabularies, contingent on a specific research setting.

KEY WORDS

consensual validation, phenomenal data, participatory sense-making, empirical phenomenology, intersubjective accessibility

CLASSIFICATION

APA: 2260, 2630

JEL: I39

*Corresponding author, η: oblak.ales.93@gmail.com; -;-

INTRODUCTION

In this article, I present a method of consensually validating *phenomenal data*. In qualitative research, consensual validation refers to the process of checking with our co-researchers whether the categories induced during the analysis of raw data correspond to their subjective reports [1]. I propose to make use of *participatory sense-making* to establish intersubjective vocabulary to describe specific aspects of our co-researchers'[1] experience. I understand phenomenal data to be consensually validated when a group of co-researchers possesses a vocabulary with which they can describe their experience.

This method of consensual validation constitutes a fusion of the *prismatic approach* and *participatory sense-making*, supported by theoretical discussions. The current iteration of the method started with the prismatic approach, a method for the study of embodied social dynamics. Based on my co-researchers' feedback, the method of consensual validation subsequently underwent a number of modifications.

The aim of this article is to offer a method of validating phenomenal data as the term is understood by Varela and Shear: subjective reports on lived experiences rather than philosophical intuitions about experience[2] [2]. While consensual validation has been claimed in various studies [3, 4], a detailed protocol has thus far not been described. Considering scientific transparency [5], in particular when it comes to qualitative research [6], explicit protocols for consensual validation are necessary.

The article is structured as follows: In the first section, I draw a comparison between phenomenology and natural history as exemplars of descriptive sciences, arguing that nominalization (i.e., naming the objects of inquiry) in both constitutes intersubjectively accessible data. This discussion works to support the position that only such data can be considered valid. In the second section, I argue that any vocabulary with which we might refer to shared experience needs to be autonomously constructed. In the third section, I present how the method was developed. In the fourth section, I present my guidelines for consensual validation of phenomenal data. In the fifth section, I discuss the epistemic status of phenomenal data validated with our approach.

A final introductory note: the method of validating phenomenal data proposed in this article is agnostic regarding the method that was used to acquire the data in the first place. It operates under a specific theory of knowledge that may or may not be admissible by some methodological frameworks in first-person research. As such, this method of consensual validation is not meant to be universal. It amounts to merely a proposal. Researchers should determine on a case-by-case basis whether the proposed method of consensual validation suits their research question and epistemological commitments.

CONSENSUAL VALIDATION AS INTERSUBJECTIVE ACCESSIBILITY: THE CASE OF PHENOMENOLOGY AND NATURAL HISTORY

In this section, I discuss the nature of intersubjectively valid data, particularly in the context of *descriptive sciences.* Descriptive sciences, broadly construed, are the sciences that outline the properties of a given phenomenon, rather than offer the conditions that give rise to said phenomenon or the mechanisms of how the phenomenon operates. What for many researchers precludes the scientific study of experience is the claim that phenomenal data are subjective, i.e., that experiential reports are inaccessible to a community of researchers, and thereby cannot constitute objects of scientific inquiry [7]. Namely, two researchers cannot observe phenomenal data as it is only accessible in the first person. I argue, however, that

through nominalization[2] phenomenal data can asymptotically approach the point where the overlap between experiences of different individuals is considerable enough to denote an intersubjectively accessible phenomenon, provided that the method of acquiring the original phenomenal data is systematic. Intersubjective accessibility would make phenomenal data accessible to a community of researchers (i.e., they would be able to agree regarding the nature of experience under investigation). This approach is valid within natural history. By analogy, I believe that this holds in descriptive sciences across the board, including phenomenology.

I will use the example of natural history to show that descriptive sciences primarily rely on nominalization, i.e., naming of phenomena under investigation, to establish intersubjective consensus about what they are studying. A common argument leveled against the scientific study of experience is that such a project can be merely descriptive [8] or that it can provide only lists of contents of consciousness [9]**.** The derision of descriptive science has been argued to be more of a prejudice [10, 11]. Biology and astronomy were both descriptive sciences [12, 13]; today, calls have been made to return descriptive work into the purview of biological sciences in general [14, 15], and descriptive approaches have become the gold standard in genetics [16]. A broader defense of descriptive sciences is beyond the scope of this article. I merely wish to suggest that many scientific disciplines rely heavily on a variety of descriptive frameworks.

I will now argue for descriptive sciences using nominalization as a method of rendering their objects of investigation intersubjectively accessible. An example of a descriptive science is biology, specifically when it assumed the form of natural history. Natural history refers to the Modern Era scientific program that aimed at establishing a taxonomy of the natural world. As Michel Foucault [17; p.114] writes in *The Order of Things*, "natural history is nothing more than the nomination of the visible." In explanation, 18th century naturalists relied on *visual epistemology* to investigate the natural world; i.e., the idea that depictions in and of themselves carry epistemic value. Practically, this means that scientists were trained in how to observe, represent, and describe the natural world. That is, the descriptions and depictions provided by experts, trained in scientific observation of the world, were ascribed a truth value. Expert descriptions of plant-life necessitated a philological approach towards the construction of knowledge, i.e., the comparison of different reports had to be conducted both at the level of observation (either of actual preserved specimens or detailed illustration) and the careful study of botanical texts. One could argue that natural history is no longer an accepted scientific program. Note, however, that genetics still constitutes a descriptive science [16].

I argue that there is significant similarity between the methods of natural history and phenomenology[3]. Much like how observation, description, comparison, and classification constituted the method of natural history (e.g., the Linnaean taxonomy of organisms [18]), phenomenology aims at classifying experience. The centrality of classification in phenomenology is summarized by Jean-Paul Sartre [19; p.5] in his account of imagination: "[P]roduce images in ourselves, reflect on these images, describe them, which is to say, try to determine and classify their distinctive characteristics."

The taxonomic streak in Sartre is self-evident, however, an important difference between phenomenology and natural history lies in the epistemic value of philology and visual epistemology. The central method of philology, i.e., the exegesis of texts, is criticized heavily by phenomenologist Paul Ricoeur [20; p.91]:

> [A]ll interpretation places the interpreter in medias res and never at the beginning nor at the end. We happen upon a conversation which has already begun and in which we try to orient ourselves in order to make our own contribution to it. But the ideal of an intuitive foundation is that of an

> interpretation which, at a certain moment, would become a total vision.

In other words, the philological method relies solely on the interpretation of texts for the construction of knowledge, while Husserlian phenomenology relies on an intuitive understanding of the subject at hand as well (although, see also [21]). Similarly, visual epistemology is somewhat problematic in phenomenology. Edmund Husserl's schematized depictions of time consciousness [22], for instance, were criticized and later on amended both by Maurice Merleau-Ponty [23] and Francisco Varela [24] (for a more recent attempt at a visual presentation of phenomenal data, see also [25]). I argue that this controversial status of visual depictions of experience stems from loss of intuitive information about the subjective dimension of the mind when reducing phenomenal data to merely the visual modality.

Now, to move from the divergences between the epistemologies of natural history and phenomenology back to their similarities. The biggest commonality between phenomenology and natural history is that they rely on a specific context of observation to achieve an understanding of their respective objects of inquiry. While knowledge construction is indeed supplemented by texts, the principal way in which naturalists obtained knowledge was through the eidetic (i.e., knowledge-giving) intuition of vision [26]. In *Visible Empire*, a monograph on 18th century botany, Daniela Bleichmar [27; p.47] writes that naturalists posited two ways of observing the world: *the spectacular gaze* and *the scientific gaze*, where

> [t]he word "spectacle" does not connote superficial entertainment: while amusing and pleasurable, nature is always instructive. […] [T]he word "spectacle" refers to a mode of seeing predicated on notions of transparency and immediacy, a way of looking that was open to everyone, regardless of background, and required no specialized training.

By contrast, scientific observation consisted of uncovering the underlying order of nature. Its "goal was not simple, immediate looking but rather expert observation, going beyond superficial traits to focus on the significant (ibid.)." The importance of specialized observation is the biggest difference between natural history and philology proper. Specifically, natural history never relied on texts as the sole source of knowledge [28]. As Bleichmar [27; p.46] writes: "Eighteenth-century natural history publications repeatedly proclaim that vision constitutes the best method for investigating nature and that images provide the preferred means of transmitting this knowledge." It is the eidetic intuition of vision that is the true bearer of knowledge in natural history rather than the texts themselves.

Phenomenology relies on a specialized form of observation as well. It observes experience while *bracketing* [An. Gr. *epoché;* Ger. *Einklammerung*] the *natural attitude.* The idea of the natural attitude is that before reflecting on it, we exist immersed in a world that quite simply appears *to be there* and we are *uncovering* it with our senses. Our understanding of the world, of the entities that inhabit it, and of our own consciousness, is laden with assumptions and theories. As Edmund Husserl [29; p.2] writes in *Thing and Space*:

> In the natural attitude of spirit, an existing world stands before our eyes, a world that extends infinitely in space that now is, previously was, and in the future will be. This world consists of an inexhaustible abundance of things, which now endure and now change, combine with one another and then again separate, exercise effects on one another and then undergo them.

However, if we are to form a theory of experience that takes into account the properties of experience, rather than imposing upon them a framework of natural sciences [30-34], we must learn how to separate ourselves from the natural attitude. We can do this by performing the act of bracketing. Bracketing refers to suspending the assumptions and theories we hold

about the mind, experience, the world and our existence in the world, including scientific theories about them [35]. It refers to an attitude of wonder before the world [36].

We must note that the spectacular gaze and the scientific gaze do not map onto the natural attitude and the act of bracketing, respectively. Instead, both the spectacular gaze and the scientific gaze exist within the natural attitude [for an empirical account, consider 37]. Indeed, in his preface to the French translation of Husserl's [38; p.xx] *Ideas Pertaining to a Pure Phenomenology and Phenomenological Philosophy*, Ricouer places the whole of natural history within the domain of the natural attitude:

> I am at first lost and forgotten in the world, lost among things, in ideas, among plants and beasts, among others […] We understand Naturalism as the lowest degree of the natural attitude and at this level it conveys its own collapse; because if I am lost in the world, I am already ready to treat myself as a thing in the world.[4]

However, the approach of natural history may guide us in constructing consensually validated phenomenal data. Specifically, if naturalism represents a way of nominalizing the visual world as seen by expert observation, phenomenology *may* be the way of nominalizing the experiential world. By adopting the assumptions of natural history, I hold that when nominalization of the experiential is done in a systematic way, it is rendered intersubjectively accessible.

The question, however, is how can we establish a vocabulary that would refer to specific aspects of experience?

PARTICIPATORY SENSE-MAKING AND AUTONOMOUS NEGOTIATION OF MEANING

The idea of establishing a shared vocabulary with which to refer to the same aspects of a given experience is not new. For example, during Heinrich Klüver's research into the structure of hallucinations, he discovered that his participants were prone to giving mystical descriptions of their experience. To circumvent this problem, he trained his participants beforehand by providing them with concepts with which to describe their experience [39].

Further, the idea of *naming* a specific aspect of one's experience is part and parcel of the qualitative approach to research, where the basic analytical tool is *coding.* We ascribe abstract descriptions to qualitative reports (typically rendered in the form of text). At the highest levels of abstraction, we may induce categories that are nothing more than words or phrases [40].

However, it is typically the researchers who name the categories under investigation and present them to their participants for them to validate the coding process[5]. From the point of view of phenomenology, this approach is problematic as what is established is not a vocabulary that could then be organically used to describe aspects of experience, not to mention that empirical phenomenological studies commonly do not even engage in the process of data validation with their participants.

Research has shown that if we wish to create a lexicon within which the meaning is grounded in the outside world, the lexemes must be autonomously constructed and negotiated by different agents in the interacting community [41]. Luc Steels attempted to solve the *symbol grounding problem* – how symbolic structures relate to the outside world [42] – by constructing artificial agents which at any time could assume control over one of two cameras placed in front of a board of colored shapes. When two artificial agents took control over the cameras, they could point to the same shape and name it. If one of the agents already had a name for the shape and the other one did not, the latter simply adopted the name. If both

agents already had a name for it, they negotiated a new name for the shape. Eventually, the community of artificial agents negotiated a shared vocabulary for all the shapes on board [41]. Steels argues that an autonomous negotiating of meaning is a necessary condition for relating a symbolic structure to an object in an outside word (i.e., for grounding a symbol).

By analogy with the symbol grounding problem, the question for empirical phenomenology is how to create a vocabulary to describe the experienced (rather than outside) world. How can we then go about solving the problem of the construction of a vocabulary with which we would be able to refer to those aspects of our experiential world that are brought to the fore by a specific research situation? One possible approach was suggested by Hanne de Jaegher [43], and Elizaveta Solomonova and Sha Xin Wei [44]: *participatory sense-making.*

Participatory sense-making refers to an enactivist theory of social cognition and language. The idea behind participatory sense-making is that within social interaction, cognizers coordinate their behaviors such that the social interaction itself becomes autonomous. By constraining each other, the cognizers modify their behavior (either bodily or linguistically) so as not to bring the interaction to an end [45, 46]. In the development of our method, the position of participatory sense-making seriously was taken seriously. It was incorporated both as a means of validating data (as will be explained below, the co-researchers have to *agree* on the framework of nominalization), as well as a criterion for validated data (i.e., the vocabulary needs to be conducive to establishing an autonomous interaction).

These theoretical discussions lead to an important methodological consideration: we do not present our co-researchers with already formed categories (although we may have provisionally constructed them during the analysis of phenomenal data). Instead, we provide them with ethnographic descriptions of experience, constructed from several reports (see below). During the validation sessions, the community of co-researchers then collectively ascribes names to these descriptions, or – if the descriptions do not correspond to their experience – either divide them further or group them together. In doing so, we further equalize the participant-researcher power dynamics: We invite our co-researchers to be part of the data analysis process rather than merely the data acquisition process.

This autonomous, collective construction of a vocabulary with which to refer to the co-researchers' experience constitutes the central innovation of this approach. Further, it represents the central criterion for whether we have managed to establish a vocabulary with which to describe our experience: if, by the end of the validation session, the co-researchers are able to discuss their experience naturally with our vocabulary without their interaction breaking down, the phenomenal data can be considered consensually validated. This principle is illustrated in Figure 1.

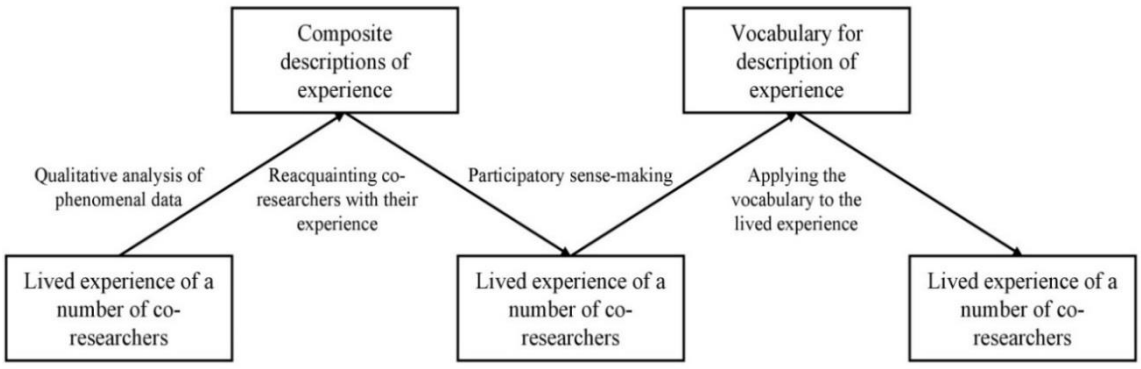

Figure 1. A schematic depiction of using participatory sense-making for consensually validating phenomenal data.

CONSTRUCTING A METHOD OF CONSENSUAL VALIDATION

In this section, I present how the method of intersubjectively validating phenomenal data was developed. I first present the starting point: the so-called *prismatic approach*. Then, I present how this approach was augmented using participatory sense-making and philosophical phenomenology. Finally, I discuss how the debriefings on individuals' experience of validating phenomenal data were used to further improve this method of consensual validation.

The method was developed alongside a longitudinal empirical phenomenological project that investigated the experience of solving a *visual span task* [47]. Throughout the remainder of the text, the reader should keep in mind that the validation sessions were conducted *after* the processes of data acquisition and analysis. As such, they do not serve as a means of gathering data. Rather, they are aimed at testing and improving the findings of the qualitative analysis. For context, I will now briefly outline the research design of the study.

RESEARCH DESIGN

The visual span task is a psychological task used to measure how many objects, presented in the visual modality, and individual can maintain in her working memory. The visual span task consisted of a presentation of a grid, in which certain cells were filled in. The grid was presented for 2 500 milliseconds. Immediately after the grid disappeared, an empty grid appeared, and our co-researchers had to fill in the black cells to match the first grid. If the co-researchers successfully reconstructed the grid, they received positive feedback, and the difficulty of the task increased. Conversely, if the co-researchers were unsuccessful in reconstructing the grid, they received negative feedback. Upon making two subsequent mistakes, the task stopped, and the co-researchers received a number denoting the span of their visual working memory. Figure 2 depicts an example of a to-be remembered grid.

The co-researchers were stopped after a random trial [48], and a phenomenological interview followed. The interview explored the time interval spanning the presentation of the stimulus, the delay, and the reconstruction of the stimulus. The interview procedure was structured based on a broadly qualitative approach. In the first part, the interview was open-ended, gathering phenomenal data based on the co-researchers' experience. Afterwards, theoretical sampling [40] was used to address specific hypotheses about the experience of a visual working memory task. All the interviews were conducted by the author of this article.

After conducting phenomenological interviews on the experience of solving the visual span task, a closed-form debriefing, based on extant discussions on the validity of phenomenal data, was used to ascertain the quality of the interview. Four interviews were conducted with each co-researcher. During each interview, only one sample of experience was gathered.

The data were analyzed according to the principles of constructivist grounded theory [40]. The study focused on two aspects of experience: the strategies used to solve a visual working memory task, and the attitudes individuals take towards the psychological task. In the analysis process, experiential categories were constructed according to these two research interests.

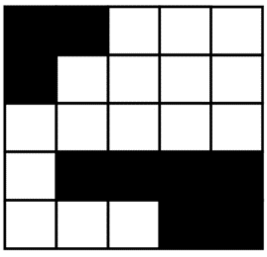

Figure 2. Example of a stimulus in the visual span task.

VALIDATING THE DATA IN THE VISUAL-SPAN STUDY

Once the phenomenal data were analyzed, consensual validation workshops informed by the prismatic approach were conducted. Consensual validation was performed in order to a) validate the analysis protocol; and b) establish an intersubjective vocabulary with which to refer to different aspects of experience of solving a visual span task.

The prismatic approach to gathering data on embodied social dynamics was developed by Barbara Pieper and Daniel Clénin [49, 50]. According to this approach, the participants of a given study attend a workshop whose goal is to investigate a specific aspect of their experience. At this workshop, the participants engage in the specific embodied social dynamic that is being investigated by jointly performing a given practice and observing how the related social dynamic unfolds. In particular, they are interested in various ways this social dynamic may or may not manifest itself.

From this workshop model suggested by the prismatic approach, the following elements were adopted:

1. The joint investigation of experience;
2. Reacquainting participants with the aspect of experience under investigation; and
3. Observing different ways in which a given aspect of experience may appear to individuals.

In line with the ideas of reflective cognitive science [51], in particular when it comes to the scientific investigation of experience [48; Ch.21], debriefings on the experience of validating the data were conducted . As several critical points entered the co-researchers' awareness, the approach to consensual validation was iteratively refined.

In total, 11 co-researchers participated in these workshops (up to four per session, spread out across four sessions). Each co-researcher (except for the principal investigator) attended only one workshop. Each workshop began with a brief lecture delivered by the principal investigator outlining the findings of the study (e.g., that we can decompose the attitude individuals take towards a visual working memory task into three dimensions), as well as the purpose of the workshop. During this initial lecture, it was emphasized that the goal is not to confirm the way the principal investigator ordered the phenomenal data during the qualitative analysis, but to challenge it. The goal of the lecture is to attempt to create as equal relationship as possible between the co-researchers. While pragmatically there is still a principal investigator guiding and organizing the study, her authoritative role should be reduced. A sense of equality among co-researchers allows for a discursive space to be opened where the co-researchers feel comfortable challenging and critiquing the current understanding of the phenomenal data.

During the validation workshops, each participating co-researcher assumed both the role of the person observing her experience and the role of interviewing the person observing her experience. Co-researchers sat together in the same room. The co-researchers took turns in solving the visual-span task, while observing their experience. *All the other co-researchers* guided the individual who performed the visual-span task through an empirical phenomenological interview. The interviews were done jointly by the whole group: that is, one person performed the visual span task while observing her experience, and all the other co-researchers interviewed her together. This process was repeated until every participating co-researcher was in the role of observing her experience.

The structure of the interview used by the group of co-researchers took the same form as during the process of data acquisition in the visual-span study: it began by a broad outline of experience, followed by pointed questions, aimed at describing each individual aspect of experience.

Further, the iterative process of observing and reporting on one's experience, and interviewing the reporting co-researchers, was repeated for every aspect of experience under investigation (i.e., one round to gather phenomenal data on strategies used to solve a visual span task, and one round to gather phenomenal data on attitudes taken towards the task).

Afterwards, the validation workshop took the form of a group discussion where we compared the way the principal investigator analyzed the data, and how the co-researchers' lived experience compares. The co-researchers were able to point out where the current state of analysis corresponds to their experience and where it deviates from it in a group setting. The last stage of the workshop was to attempt to jointly construct mutually agreed upon names for the aspects of experience under investigation.

After the validation session, the co-researchers were given the following semi-structured questionnaire to report on their experience of the validation session itself:

1. Do you believe we succeeded in establishing a common language to describe the experience of solving a visual working memory task?
2. Do you believe that the group setting changed your experience or your reporting on the experience?
3. Do you believe that aspects of experience that you reported on during the group discussion represent your experience during the performance of the visual working memory task?
4. Do you believe that having to report on your experience in language changed your experience?
5. What do you think could be changed in our approach to consensual validation?

Based on the responses to these questions, the structure of the workshop was iteratively modified: the subsequent workshop took into account the feedback from the previous one and was adapted accordingly. In the next section, the guidelines towards consensually validating phenomenal data by using participatory sense-making are presented.

PARTICIPATORY SENSE-MAKING AS A METHOD OF VALIDATING PHENOMENAL DATA: THE GUIDELINES

In the present section, the method of validating phenomenal data is presented. Crucially, the validation workshops are conducted *after* the process of data acquisition is complete and on the *same* sample of co-researchers. Based on the theoretical principles outlined in the first two sections, and the feedback gained from our workshops, the following guidelines towards using participatory sense-making as a method of consensually validating phenomenal data were reached.

1. The starting point are not categories constructed during the qualitative analysis of the data, but composite ethnographic descriptions;
2. (If possible) the experiences under investigation are provoked during the validation session;
3. Co-researchers are asked to observe whether it is necessary that their experience corresponds to the composite descriptions or whether it can deviate;
4. Co-researchers collectively construct the vocabulary with which to refer to their experience;
5. The co-researchers once again observe their experience (if that is possible), and describe it with the newly-established vocabulary.

In this section, we will take a closer look at each of these steps, providing principled and empirical support for them.

COMPOSITE DESCRIPTIONS OF EXPERIENCE

The starting point is phenomenal data that has been subjected to some form of analysis. At present, this method of consensual validation is not committed to adhering either to approaches to data analysis that are tailor-made for experiential reports [52, 53] or qualitative

data in general [40, 54]. Rather, in light of qualitative research's commitment to methodological pluralism [55]**,** I consider various approaches to data analysis equally valid.

Importantly, we do not set out to validate named categories as established by the qualitative analysis of the phenomenal data. Rather, we present our co-researchers with composite ethnographic descriptions. Composite descriptions refer to a *style of presentation* whereby we do not present concrete descriptions of the phenomenon under investigation, but we combine a number of descriptions and abstract them away into a typical and telling example [56].

The principal investigator presents these composite descriptions as well as a detailed research design during an introductory lecture. The purpose of this lecture is to inform the co-researchers of all the aspects of the study (which may have been unclear to them during the initial interviews), as well as to explicitly let them know that there is nothing that is being hidden from themThis position of empowerment is crucial so as to minimize the influence of demand characteristics [57] on their reports. Specifically, we do not wish for them to attempt to guess what the goal of the research is [58]. Rather, we explicitly inform them about the research goal, and then challenge them to attempt to challenge our categories. This challenge is both stated during the introductory lecture and made a part of the process of going through experiential categories later on[6].

It is important that the overall arc of the workshop is clear to the co-researchers. As one co-researcher reports in her feedback on the workshop:

> Because we were mostly talking, I feel that everything was left hanging in the air. We never reached any clear conclusions. We could write down different ways in which others experienced things and then through conversation see which experiences overlap or are related to each other. (Co-researcher 14, *feedback on the workshop*)

The feedback that the discussions were floating in the middle of nowhere was shared by many co-researchers, in particular those who were involved in longer sessions. I suggest that ahead of the validation session, the co-researchers be given all the composite descriptions with room allocated for notes and eventual names. This allows them to make notes on aspects of experience not currently under investigation, as well as give them a sense of completion.

REACQUANTING THE CO-RESEARCHERS WITH THE EXPERIENCE UNDER INVESTIGATION

Once we make the aim of the study known, we invite the co-researchers to reacquaint themselves with the experience under investigation. If this experience can be easily induced (e.g., by means of a psychological task), each co-researcher is given the possibility of privacy to observe their experience once again (i.e., we provoke it) [3, 52, 53]. If, on the other hand, the experience cannot be reproduced, the co-researchers are given time to consult their journals. Importantly, reacquainting with the experience should be done in a setting that reflects the social context of the original investigation, as was reported by one of our co-researchers:

> We should not do the task together during the group session (with people watching what you are doing) because this aspect changed my experience quite a lot and was pretty distracting. I felt less secure about my report than when I did the task on my own during the [original sessions], and my experience was quite a bit different back then. (Co-researcher 11, *feedback on the workshop*)

JOINT INVESTIGATION OF EXPERIENCE

Then, the co-researchers interview a single individual on her experience. We continue with this process until every co-researcher has been in the role of the reporting on their experience. In other words, individuals who had previously simply participated in the study (by observing and reporting on their experience) now help guide each other towards observing and describing their experience (i.e., they become interviewers). During the process of mutual interviewing, the co-researchers are invited to again observe their experience or to read up on their journal notes. They are prompted to observe specific qualities of their experience:

1. is the experience under investigation necessarily such as it was observed during the qualitative analysis, or can it be modulated with specific mental gestures?
2. can you observe specific differences between two closely related experiential categories?

This line of questioning (i.e., observing whether a specific aspect of experience is necessarily structured in such a way as observed during the acquisition of phenomenal data or if it can differ in some respect) may be problematic within schools of thought in empirical phenomenology that emphasize open-beginninged attitude towards gathering data [3, 48]. As will be explained in Section 5, the presented method for validation of phenomenal data adheres to process-oriented constructivism. Namely, we are trying to establish a framework that is contingent (i.e., there are potentially infinite possible ways of dividing the categories), and simultaneously internally consistent. In this case, an example of an internal inconsistency would be two categories that would describe the same aspect of one's experience. A similar approach was adopted by Husserl [35; p.220] (emphasis in the original):

> Since every negatum and affirmatum is itself an Object posited as existent, it can, like everything intended to as having a mode of being, become affirmed or denied. *In consequence of* the constitution of something as existent effected anew at every step, *an ideally infinite chain of reiterated modifications* therefore results.

Or, on Andrea Staiti's [59; p.815] paraphrase: "Higher-order affirmation or negation occurs when we set out to revisit a foregoing simple judgment in order to confirm or disconfirm the veridicality of its proposed state of affairs." In explanation, we check whether the structure of a specific aspect of experience – as observed in the acquisition of phenomenal data (whether it be through philosophical intuition or second-person methods) – reflects the lived experience of our individuals, we attempt to provoke experiences that either conform to or deviate from it.

During the joint phenomenological interview, our goal is to contrast different experiences of the same phenomenon, trying to ascertain which aspects of experience necessarily remain the same across individual co-researchers, and which aspects may vary. These variations in experience need not be explicit. As the following co-researcher reports, just witnessing how other people experience the same experimental setting, may prompt them to observe their own experience differently:

> My experience changed in that I started to pay attention to new aspects of experience that were described by the others but that I was not aware of before. I tried to check if I had a similar experience as them, that I just had not realized before, or if these aspects were completely absent to me. But actually, I do not think that my experience changed by that but rather the awareness of my experiences. (Co-researcher 11, *feedback on the workshop*)

I will now demonstrate two examples of the attempt at challenging the induced experiential categories. To reiterate: these categories were first induced by the principal investigator during the qualitative analysis stage of the study. In the validation workshops, the

co-researchers are presented with a composite description of the experiential category. They attempt to discover whether this experiential category is something that they can observe in their experience or if it deviates from it.

The first example of the validation interview during the workshops represents a case that led to an elimination of the category *experimental orientation* – how it feels to be a participant in a psychological experiment. Importantly, during the initial process of data acquisition, several reports were gathered on how one's experience changes when participating in an experiment:

> Co-researcher 7: Certainly, [being in an experiment] feels different. It cannot be compared to anything experienced outside of this setting. So, trying to think of the things I do inside of this setting, of the skills you need to memorize it, whether it's committing a poem to memory, it's different, because you are not trying to compare yourself to others. And then also, sort of the complete setup, of somebody watching you. It was not that I was thinking *I have to prove myself, I have to solve 75 of them,* but it certainly alters how you engage with it. And also what you perceive and notice. (Co-researcher 7, *validation interview*)

> Co-researcher 9: I just noticed another really important dimension about how it feels to be me in an experiment. It's this dimension of how much you are trying to understand the task itself behind everything. So, going through one pattern, and another pattern, it is one thing. But I often simultaneously try to understand how the task itself works. I mentioned that in our Interview. I noticed that it always starts with four squares, and every time you get two of them correct, the number increases by one. And in this way, you can predict how many black squares in total will appear in the next square, and I was using this knowledge to check if I have all of them. I counted them and I knew how many of them there had to be. (Co-researcher 9, *validation interview*)

The consensual validation revealed that we may break this experience down into its various social dimensions, as well as the attitude of being goal oriented. While these two aspects of experience are prevalent when somebody is undergoing an experimental situation, they are not unique to it. Consider the following example:

> Co-researcher 7: It sometimes feels in a museum that I have this *divide-and-conquer feeling.* How I look at the image and parts of the image, and I try to understand what it shows, what it means. That can be very prominent as a strategy that unfolds itself. So, sometimes going to the museum can be quite exhausting for me and sometimes it does not feel like I am going there to enjoy the paintings but that I am going there to look and understand. That sort of creates this *task-mentality*. (Co-researcher 7, *validation interview*)

Prior to the validation study, a category *experimental orientation* was induced during qualitative analysis. *Experimental orientation* was an experiential category that described the particular atmosphere of experience that is present when one is interacting with a psychological task in a research setting. Experimental *orientation* satisfied many of the standard criteria in qualitative research for it to be induced as a specific category:

1. it was reported by a number of different co-researchers;
2. it was clearly distinguishable from other categories, and we observed a number of limiting cases (i.e., clear instances of phenomenal data that did not conform to the category of *experimental orientation*).

However, during the validation session it became apparent that while the experience of *solving a visual-spatial working memory task* was a unique experience within the context of this study, the experience of participating *in a study* was not. This led me to abandon it as an experiential category, as it amounted to an interpretation on the part of the principal investigator, rather than a faithful reflection of the lived experience of the co-researchers.

For the second example, let us look at how we constructed a new category during validation sessions. During the qualitative analysis of the data, two categories related to association were induced. The first was an

> *experience in which a to-be-remembered stimulus is accompanied by a visual feeling of something that it resembles. This visual feeling appears as a mental image that is either projected in the outside world, or exists in an internal mental space.* The second was an *experience in which a to-be-remembered stimulus is accompanied by a clear idea of what it resembles. This idea may be explicitly articulated in an inner voice.*

This potential category may be demonstrated with a picturesque example observed in the study. A co-researcher was reminded of the swastika by the to-be-remembered stimulus. As I will demonstrate the differences between the established categories visually, I will replace the swastika with the symbol of the fictional state of Tomania from the film *The Great Dictator* [60]. In the co-researcher's experience, this association was accompanied by mental imagery. Ultimately, this aspect of experience was named "visual image."

Another co-researcher experienced a sense of the to-be-remembered stimulus being like something that might be used as a symbol in a totalitarian regime. As he noticed that this configuration of black-and-white squares was fairly common, he ascribed the meaning of totalitarian iconography to it. This aspect of experience was ultimately named "symbolized description." These two aspects of experience are depicted in Figure 3 as a) and b).

During the validation session, however, we concluded that visual image corresponds to two aspects of experience: one that is comprised by the experience of imagery, and the other comprised of various existential feelings [61]. Thus, we introduced a new category: *atmospheric image*. In Figure 3, the new categories are depicted under headings c) and d). Such a refinement and removal of categories is the central goal of the joint interview: we wish to find aspects of experience that are stable and understandable across co-researchers.

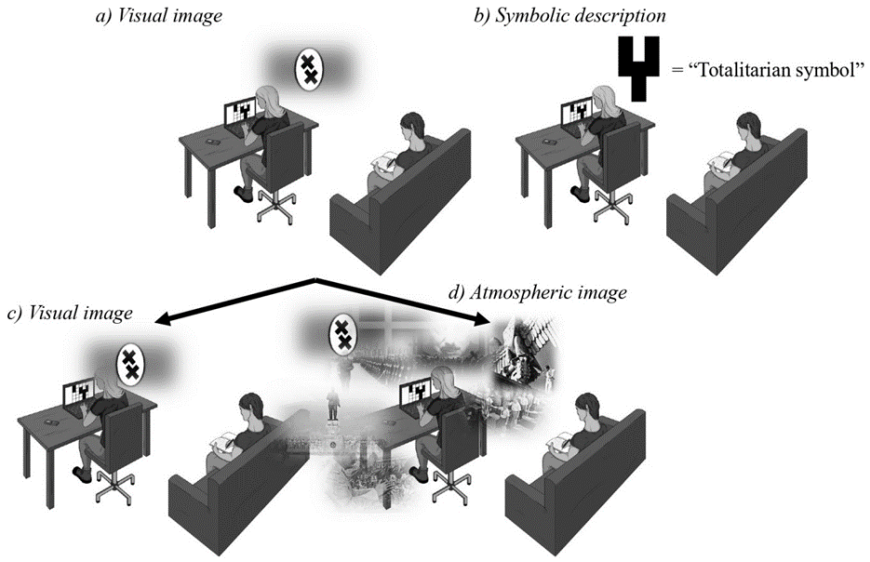

Figure 3. Categories associated with association.

NOMINALIZING EXPERIENTIAL WORLDS

After all the co-researchers were both in the role of observing and reporting on their experience, *and* guiding others through a second-person phenomenological interview, we collectively decide on how to name a specific aspect of experience. For example, consider the following exchange between four co-researchers:

> Co-researcher 11: The only thing that feels forced to me is to put competitiveness between me and other people, and against myself on the same axis. I do not know.
>
> Principal investigator: Would you say that these are two different things?
>
> Co-researcher 11: Yeah, it's a very different feeling.
>
> [...]
>
> Co-researcher 10: Exactly! There's an element of jealousy when it's competitiveness with other people.
>
> Co-researcher 11: Yes! Yes!
>
> Co-researcher 10: There's this element that you want to be better than other people
>
> Co-researcher 9: Yeah, competitiveness with me is something that I never experienced as negative. It was more encouraging. So, yeah, I would call it ambition. (validation workshop)

It is important that each individual is comfortable with the names, and any concerns regarding the vocabulary should be addressed. The final step in the process of validation is for co-researchers to once more be exposed to the experience under investigation, either by provoking it or by consulting their journals. In this step, the co-researchers must be able to discuss their experience with one another with our vocabulary (as per participatory sense-making, the interaction must not break down). Being able to organically use the vocabulary constitutes the most important criterion for an intersubjectively accessible vocabulary on experience. In the study, this was achieved during the third validation workshop, where co-researchers were able to discuss their experience of solving a visual-span task in great detail using the jointly agreed upon vocabulary.

THE EPISTEMIC STATUS OF CONSENSUALLY VALIDATED PHENOMENAL DATA

In the previous section, I laid out the proposed method of consensually validating phenomenal data. Importantly, my approach hinges on a group of co-researchers observing their experience, comparing said experience to composite descriptions of similar experience, and then jointly constructing a vocabulary with which to describe them.

This method of consensual validation operates under the *constructivist epistemology*. Constructivism claims that knowledge is not uncovered from an observer-independent world, but rather that it is constructed by the researchers [62, 63]. In relation to consensual validation, this means that we as co-researchers do not agree on what objectively exists in our experience, but rather that we collectively create a contingent body of knowledge about our experience. It is not knowledge about *experience as is,* but *experiential reports as constructed in a given study* [34]. This epistemology can be specified further: the proposed method of consensual validation follows constructivist epistemology augmented by *process-oriented ontology* [64], developed by Alfred North Whitehead. In *Process and Reality*, Whitehead [65; p.67] writes:

> Actual entities atomize the extensive continuum. This continuum is in itself

> merely the potentiality for division; an actual entity effects this division. The objectification of the contemporary world merely expresses that world in terms of its potentiality for subdivision and in terms of the mutual perspectives which any such subdivision will bring into real effectiveness.

In explanation, the processes that constitute the world can be divided in any number of ways. How we divide the world, however, is entirely open-ended. The potentiality of the world may therefore contain contradictions, but once we make it concrete and discrete, the world is divided into an internally coherent system. In Whitehead's process-oriented philosophy, entities refer to parts of the world that are spatially and temporally located. However, as will be seen below, this framework has been productively used to analyze experiential worlds as well [64].

Potentiality for division could consistently be observed in the visual-span study as the frameworks of experiential categories that were constructed did not apply in their entirety for all of the co-researchers. What applied to an individual co-researchers was some subset of this framework. Consider the following report on the validation session:

> We found multiple common languages. Or rather, I feel that we reached two or three common ways of experiencing the task, and mostly each person could identify themselves with (at least) one way. (Co-researcher 14, *feedback on the workshops*)

The result of this approach is not the construction of *the* language for describing experience, but the construction of *a* language. In the following quote from a co-researcher, we can see that this language is approximately precise, which means that it is both adequate for the description of our experience, while it simultaneously remains possible to articulate it in a slightly different way:

> Sometimes I felt just a tiny little bit forced to fit my experiences into given categories and dimensions because I thought it would reduce them a bit too much. On the other hand, they did fit into the suggested categories overall, and I imagine that some sort of reduction might be necessary to find common patterns or reach common grounds. It's a bit as if you would measure many people's heights and some would be 1,745 m and some would be 1,748 m and you put them both into the category of being 1,74 m. It's some sort of reduction of data, but I do not think that these 0,003 m make that much of a difference in our world. (Co-researcher 11, *feedback on the workshops*)

Figure 4. Schematic depiction of fitting our vocabulary to experience.

This idea of the method yielding more than one vocabulary for the description of experience is illustrated in the Figure 4. Imagine that we have established a vocabulary to describe the experience as it is given to two co-researchers, John and Mary. We may claim that John and Mary's experiences (as it is given to them, respectively) only partially overlap. Their respective experiences approximately map onto the established vocabulary. We can see that there are subjectively judged lacunae both in how adequate our vocabulary is in describing John and Mary's experience as well as in how many aspects of experience are shared between John and Mary. Therefore, our vocabulary is necessarily only approximately accurate in describing our co-researchers' experience.

To understand what this idea of potentiality means for the scientific study of experience, we can defer to Siegfried Schmidt. He introduces the idea of *positings*, the assumptions that inform (and by extension construct) our experiential worlds. "Whatever we do," writes Schmidt [64; p.43], "we do in the Gestalt of a positing: we do this, and not something else, although we could have done that." These positings are ways of constructing our experiential worlds. We attain positings both through our previous experience as well as through the culture that we are immersed in. Our experiential world therefore consists of what we expect it to consist of based on our prior experience. It is contingent – but it is not arbitrary.

One of the most important ways in which reality is constrained (and therefore made non-arbitrary) is through the social dimension. As Schmidt writes:

> The coupling of process results and their attribution as "real for..." must be socially accepted and thus intersubjectively confirmed, i.e., without the others there is neither certainly nor uncertainly for us. This means that experiencing something as real presupposes the context of acting and communicating communities determined by their framework of interactive dependencies [...] We necessarily live out our life-worlds together with other people [64; p.4].

In the scientific study of experience, we are "measuring" precisely observer-dependent worlds, as the observers are the only experts in their experience, as well as the only instrument of measurement through which their experience can be made accessible. If we – as a community – "deem real is real in its *consequence* [64; p.6]" at least to the extent as it appears to us in our experience, we can add another goal to the process of consensual validation. Not only do we check whether our qualitative analysis of the data was consistent with our co-researchers' experiences, but we can also jointly establish what the meaningful experiential qualities are and thereby solidify them in their experience. I observed exactly such a phenomenon in the visual-span study:

> Principal investigator: Would you say that this category that I just suggested to you matches your experience, this sense of impression?
>
> Co-researcher 6: If I look at the screen now, I can see it. I can experience this raw feeling. I can experience it if you give it to me, but before that, no. There's nothing there before that. (Co-researcher 6, *Data acquisition*)

The experience may have always been there (perhaps as an element of pre-reflective consciousness, [66]) or we may have elicited it in her through suggestion [67]. Whatever may be the case, now that the experiencer has access to a word with which to describe it, that aspect of her experience appears more salient to her.

Let us again defer to the hypothetical case of John and Mary's experience. During the moment Mary genuinely lived her experience she was more or less aware of the aspects of her experience seen in light-to-mid gray (Figure 5, left). She paid little attention to the dark

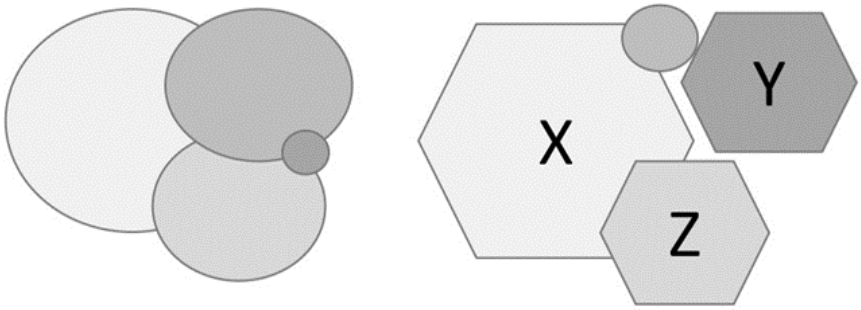

Figure 5. Mary's experience before (left) and after (right) consensual validation.

gray aspect of her experience. It still constituted *how it was to be Mary* in that moment, however, she did not bring it to the forefront of her awareness. Then, we conduct consensual validation, establishing the language with which to describe experience consisting of elements {X, Y, Z}. Not only have we now divided Mary's continuous field of experience into defined elements, we have also made them more salient in her awareness. The dark gray element, now nameable as Y, has become more prevalent in her experience. On the other hand, we failed to name the mid-gray element, which is not encompassed in our language. As Mary is now more poised to observe X, Y, Z, the salience of the mid-gray element has been reduced (Figure 5, right).

To reiterate, the epistemic value of the phenomenal data constructed through the proposed method of consensual validation relates to constructivist epistemology in two ways. Firstly, because the continuous flow of our experience can be divided and categorized in ways that are contingent, i.e., can vary across different individuals describing a similar aspect of their experience, we are not establishing the only framework with which to describe a particular aspect of our experience, but merely one of the possible frameworks. Secondly, the framework of experiential categories that we set-up during the participatory sense-making session then works to constrain our co-researchers' experience in the future.

CONCLUSION

I have offered a method of consensually validating phenomenal data. I constructed the method based on similarities between phenomenology and other descriptive sciences, exemplified by 18th century natural history. While phenomenology and natural history both constitute descriptive sciences, whereby their task is to observe, describe, and classify phenomena under investigation, there are some major differences between them. The foremost difference between the two disciplines is that natural history relies on visual epistemology and exegesis of texts. Neither of these approaches is tenable within phenomenology. An important commonality between the disciplines, however, is the importance of intuitive, expert observation. As natural history divides the ways of observing the world into the spectacular and the scientific view, so does phenomenology divide its observation into the natural attitude and bracketing thereof. While these two reductions do not map one onto the other, they still attest to the intuitive way knowledge is given to observers. Based on this kinship, I argued that much like the nominalization of the visual world led to an intersubjectively observable body of scientific knowledge in natural history, nominalization of experiential world may lead to intersubjectively accessible phenomenal data.

I suggested that the nominalization of the experiential world should be done by an autonomous group of co-researchers, rather than being imposed by the principal investigator and checked against the experience of the co-researchers. I derived the importance of

autonomous construction of vocabulary both from discussions on the symbol grounding problem, and participatory sense-making. To implement this solution, I iteratively modified the prismatic approach, a method of gathering data on social dynamics. Based on reports on experience of consensual validation, I modified the method.

The final iteration of my approach consists of a group of co-researchers being exposed to composite descriptions of experience. They are invited to once again observe their experience, comparing it to the composite descriptions. They are asked to attempt to falsify the composite descriptions. Finally, co-researchers construct a possible vocabulary for description of experience under investigation. This vocabulary is contingent on a particular research setting, but nonetheless offers a way of operationalizing aspects of experience.

REMARKS

[1]In line with empirical phenomenology, I use the term *co-researcher* instead of the traditional term *participant* [48, 68, 69].

[2]The discussion about the subjective nature of data on experience is broad and beyond the scope of this article. For example, Hurlburt [48] claims that subjective reports gathered with the descriptive experience sampling technique are *radically non-subjective*, in the sense that there is a correct way of describing what an individual was experiencing in a given moment. Here, I wish to address what Froese and colleagues [70] would refer to as the *theory-experience gap:* the difference between what a researcher might theorize individuals will experience in a given context and what the individuals end up experiencing in that context. We may introduce a third type of experience: naive subjective reports, which are the beliefs individuals may have about their experience. Among these three types, only systematically acquired data on lived experience (the second type) may be considered to be intersubjectively accessible.

[3]*Nominalism* is a philosophically burdened term. As Ian Hacking [71; p.81] writes in *The Social Construction of What?,* nominalists hold that "[t]he world is so autonomous, so much to itself, that it does not even have what we call structure in itself. We make our puny representations of this world, but all the structure of which we can conceive lies within our representations." Hacking writes on: "The nominalist retorts that we have a good deal to do with organizing what we call a fact. The world of nature does not just come with a totality of facts: rather it is we who organize the world into facts [71; p.174]." In other words, nominalism does not claim that by naming the world we "cut it at the joint," but rather that we impose structure upon it. As will be seen in Section 5, I am sympathetic to this view, i.e., when we name experience, we do not name it according to the underlying structure of experience. Rather, by naming experience, we impose structure upon it.

[4]I use the term *phenomenology* in a somewhat monolithic sense, disregarding the many different approaches within phenomenology, such as *empirical phenomenology* [3, 48, 69, 72], philosophical *phenomenology* [22, 29, 35] and *phenomenological* psychopathology [73] to name just a few. While these different schools of thought have different epistemological commitments, I believe that our method may be useful in bringing descriptions of experience to the fore of a community of researchers, regardless of whether the original phenomenal data was obtained through philosophical intuition or various second-person methods.

[7]I thank Clémence Compain for help with the French translation.

[6]I do not wish to suggest that this one-directional nature of qualitative research has never been addressed. Indeed, one of the cornerstones of the *constructivist grounded theory* approach to qualitative analysis is to account the ways in which the researcher *constructs* rather than *discovers* knowledge [63, 74, 75].

[7]A principled criticism of this explication of phenomena under investigation is that we may induce the experience through suggestion effect. Pete Lush and colleagues [67] have observed that in the general population the rate of suggestibility of individuals is normally distributed, i.e., among highly suggestible individuals, instructions in a research setting may indeed cause them to have the experience presupposed by the research design. While this effect is problematic for empirical research in mind sciences for a number of reasons, we believe that as long as it indeed *provokes* an experience in our co-researchers rather than merely lead them to say they experienced something, this does not reduce the epistemic value of our phenomenal data.

ACKNOWLEDGMENTS

The author wishes to thank Ema Demšar and Damar Hoogland for help with preparing this manuscript.

REFERENCES

[1] Eisner, E.W.: *The Enlightened Eye: Qualitative Inquiry and the Enhancement of Educational Practice.*
Macmillan, New York, 1991,

[2] Varela, F.J. and Shear, J.: *First-person Methodologies: What, Why, How?*
Journal of Consciousness Studies **6**(1), 1-4, 1999,

[3] Petitmengin, C., et al.: *What is it like to meditate? Methods and issues for a microphenomenological description of meditative experience*.
Journal of Consciousness Studies **23**(5-6), 170-198, 2017,

[4] Kordeš, U.; Oblak, A.; Smrdu, M. and Demšar, E.: *Ethnography of Meditation: An Account of Pursuing Meditative Practice as a Tool for Researching Consciousness.*
Journal of Consciousness Studies **26**(7-8), 184-237, 2019,

[5] Nosek, B.A., et al.: *Promoting an Open Research Culture: Author Guidelines for Journals Could Help to Promote Transparency, Openness, and Reproducibility*.
Science **348**(6242), 1422-1425, 2015,
http://dx.doi.org/10.1126/science.aab2374,

[6] Moravcsik, A.: *Transparency: The Revolution in Qualitative Research.*
Political Science & Politics **47**(1), 48-53, 2014,
http://dx.doi.org/10.1017/S1049096513001789,

[7] Hurlburt, R.T. and Schwitzgebel, E.: *Describing Inner Experience? Proponent Meets Skeptic.*
The MIT Press, Cambridge, 2007,

[8] Dennet, D.: *Consciousness Explained.*
Back Bay Books, New York, 1992,

[9] Engelbert, M. and Carruthers, P.; *Descriptive Experience Sampling: What is It Good For?*
Journal of Consciousness Studies **18**(1), 130-149, 2011,

[10] Grimaldi, D.A. and Engel, M.S.: *Why Descriptive Science Still Matters.*
BioScience **57**(1), 646-647, 2007,
http://dx.doi.org/10.1641/B570802,

[11] Casadevall, A. and Ferric C.F.: *Descriptive Science.*
Infection & Immunity **76**(9), 3835-3836, 2008,
http://dx.doi.org/10.1128/IAI.00743-08,

[12] Haeckel, E.H.P.A.: *Art forms of nature. In German.*
Bibliographisches Institut, Vienna, 1904,

[13] van Loon, F.W.: *Planisphaeri Celeste*.
http://upload.wikimedia.org/wikipedia/commons/7/75/Planisph%C3%A6ri_c%C5%93leste.jpg, accessed 26th December 2019,

[14] Dupré, J.: *In Defense of Classification.*
Studies in History and Philosophy of Biological and Biomedical Sciences **32**(2), 203-219, 2001, http://dx.doi.org/10.1016/S1369-8486(01)00003-6,
[15] Dupré, J.: *A Process Ontology for Biology.*
Physiology News **100**(1), 33-34, 2015, http://doi.org/10.36866/pn.100.33,
[16] Yarkoni, A.: *The Generalizability Crisis.*
PsyArXiv, 2019, http://dx.doi.org/10.31234/osf.io/jqw35,
[17] Foucault, M.: *The Order of Things: An Archeology of the Human Scieneces.*
Routledge, London, 2002,
[18] Linnaeus, C.: *The system of nature is divided into three kingdoms of nature: based on classes, orders, genera, species with characters and differences.* In Latin.
http://www2.linnaeus.uu.se/online/animal/1_1.html, accessed 26th December 2019,
[19] Sartre, J.: *The Imaginary: A Phenomenological Psychology of the Imagination*.
Routledge, New York, 2010,
[20] Ricoeur, P.: *Phenomenology and Hermeneutics.*
Noûs **9**(1), 85-102, 1975, http://dx.doi.org/10.2307/2214343,
[21] Giorgi, A.: *The Descriptive Phenomenological Method in Psychology.*
Duquesne University Press, Pittsburgh, 2009,
[22] Husserl, E.: *On the Phenomenology of Consciousness of Internal Time (1893-1917).*
Kluwer Academic Publishing, Dordrecht, 1991,
[23] Merleu-Ponty, M.: *Phenomenology of Perception.*
Routledge, London, 2012,
[24] Varela, F.J.: *The Specious Present: A Neurophenomenology of Time Consciousness.*
In: Petitot, J.; Varela, F.J.; Pachoud, B. and Roy, J., eds.: *Naturalizing Phenomenology: Issues in Contemporary Phenomenology and Cognitive Science*. MIT Press, Cambridge, 1999,
[25] Haun, A. and Tononi, G.; *Why Does Space Feel the Way It Does? Towards a Principled Account of Spatial Experience.*
Entropy **21**(12), 1160, 2019, http://dx.doi.org/10.3390/e21121160,
[26] Atran, S.; *Cognitive Foundations of Natural History.*
Cambridge University Press, Cambridge, 1993,
[27] Bleichmar, D.: *The Visible Empire: Botanical Expeditions and Visual Culture in the Hispanic Enlightenment.*
The University of Chicago Press, Chicago, 2012,
[28] Derrida, J.: *Of Grammatology.*
Johns Hopkins University Press, Baltimore, 1998,
[29] Husserl, E.: *Thing and Space: Lectures of 1907.*
Kluwer Academic Publishings, Dordrecht, 1997,
[30] Varela, F.J.; Thomspon, E. and Rosch, E.: *The Embodied Mind: Cognitive Science and Human Experience.*
MIT Press, Cambridge, 2017,
[31] Flanagan, O.J.: *Consciousness Reconsidered.*
MIT Press, Cambridge, 1992,
[32] Roy, J., et al.: *Beyond the Gap: An Introduction to Naturalizing Phenomenology.*
In: Petitot, J.; Varela, F.J.; Pachoud, B. and Roy, J., eds.: *Naturalizing Phenomenology: Issues in Contemporary Phenomenology and Cognitive Science*. MIT Press, Cambridge, 1999,
[33] Thompson, E.: *Mind in Life: Biology, Phenomenology, and the Sciences of the Mind.*
Harvard University Press, Cambridge, 2007,

[34] Kordeš, U. and Demšar, E.: *Towards the Epistemology of the Non-Trivial: Research Characteristics Connecting Quantum Mechanics and First-person Inquiry.*
Foundations of Science **26**(1), 187-216, 2019,
http://dx.doi.org/10.1007/s10699-019-09638-z,

[35] Husserl, E.: *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy.*
Martinus Nijhoff Publishing, The Hague, 1983,

[36] Fink, E.: *The* Phenomenological *Philosophy of Edmund Husserl and Contemporary Criticism.*
In: Elveton, R.O., ed.: *The Phenomenology of Husserl: Selected Critical Readings*. Quadrangle Books, Chicago, 1970,

[37] Albertazzi, L.; Reconsidering *Morphology Through an Experimental Case Study.*
Biological Theory **12**(3), 131-141, 2017,

[38] Husserl, E.: Idées directrices *pour une phénoménologie.*
Gallimard, Paris, 1950,

[39] Klüver, H.: Mescal and *the Mechanisms of Hallucination.*
The University of Chicago Press, Chicago, 1966,

[40] Charmaz, K.: Constructing *Grounded Theory: A Practical Guide Through Qualitative Analysis.*
Sage Publications, London, 2004,

[41] Steels, L.: *The Symbol Grounding Problem Has Been Solved. So What's Next?*
In: de Vega, M.; Glenberg, A. and Graesser, A., eds.: *Symbols and Embodiment: Debates on Meaning and Cognition*. Oxford University Press, Oxford, 2008,

[42] Harnad, S.: *The Symbol Grounding Problem.*
Physica **42**(1), 335-346, 1999,
http://dx.doi.org/10.1016/0167-2789(90)90087-6,

[43] De Jaegher, H.: *Intersubjectivity in the Study of Experience*.
Constructivist Foundations **14**(2), 393-395, 2016,

[44] Solomonova, E. and Wei, S.X.: *Exploring the Depth of Dream Experience: The Enactive Framework and Methods for Neurophenomenological Research.*
Constructivist Foundations **11**(2), 407-442, 2014,

[45] De Jaegher, H. and Di Paolo, E.: *Participatory Sense-making.*
Phenomenology and the Cognitive Sciences **6**(4), 485-507, 2007,
http://dx.doi.org/10.1007/s11097-007-9076-9,

[46] Cuffari, E.; De Jaegher, H. and Di Paolo, E.: *From Participatory Sense-making to Language: There and Back Again.*
Phenomenology and the Cognitive Sciences **14**(4), 1089-1125, 2015,
http://dx.doi.org/10.1007/s11097-014-9404-9,

[47] Repovš, G. and Baddeley, A.D.: *The multi-component model of working memory: Explorations in experimental cognitive psychology.*
Neuroscience **139**(1), 5-21, 2006,
http://dx.doi.org/10.1016/j.neuroscience.2005.12.061,

[48] Hurlburt, R.T.: *Investigating Pristine Inner Experience: Moments of Truth.*
Cambridge University Press, Cambridge, 2011,

[49] Pieper, D. and Clénin, D.: *Verkörperte Selbst- und Fremdwahrnehmung sozialen Handelns. Eine praktisch-theoretische Forschungsperspektive*. In German.
In: Boehle, F. and Weihrich, M., eds.: *Die Körperlichkeit sozialen Handelns. Soziale Ordnung jenseits von Normen und Institutionen*. Transcript, Bielefeld, 2010,
http://dx.doi.org/10.1515/9783839413098,

[50] De Jaegher, H.; Pieper, B.; Clénin, D. and Fuchs, T.: *Grasping Intersubjectivity – An Invitation to Embody Social Interaction Research.*
Phenomenology and the Cognitive Sciences **16**(3), 491-523, 2017,
http://dx.doi.org/10.1007/s11097-016-9469-8,

[51] Demšar, E.: *The circular character of the conceptual space of cognitive science: between scientific and lived realities of the mind.* M.Sc. Thesis.
University of Ljubljana, Ljubljana, 2017,
[52] Petitmengin, C.; Remillieux, A. and Valenzuela-Moguillansky, C.: *Discovering the Structures of Lived Experience: Towards a Micro-phenomenological Analysis Method.*
Phenomenology and the Cognitive Sciences **18**(4), 691-730, 2018,
http://dx.doi.org/10.1007/s11097-018-9597-4,
[53] Valenzuela-Moguillansky, C. and Vásquez-Rosati, A.: *An analysis procedure for the micro-phenomenological interview.*
Constructivist Foundations **14**(2), 123-145, 2019,
[54] Flick, U.: *An Introduction to Qualitative Research.*
Sage, London, 2009,
[55] Frost, N., et al.: *Pluralism in Qualitative Research: The Impact of Different Researchers and Qualitative Approaches on the Analysis of Qualitative Data*.
Qualitative Research **10**(4), 441-460, 2010,
http://dx.doi.org/10.1177/1468794110366802,
[56] Murchinson, J.M.: *Ethnography Essentials: Designing, Conducting, and Presenting your Research.*
Jossey-Bass, San Francisco, 2010,
[57] Orne, M.: *On the Social Psychology of the Psychological Experiment: With Particular Reference to Demand Characteristics and their Implications*.
American Psychologist **17**(11), 776-783, 1962,
http://dx.doi.org/10.1037/h0043424,
[58] Nichols, A.L. and Manner, J.K.: *The Good Subject Effect: Investigating Participant Demand Characteristics.*
The Journal of General Psychology **135**(2), 151-166, 2008,
http://dx.doi.org/10.3200/GENP.135.2.151-166,
[59] Staiti, A.: *Husserl and Rickert on the Nature of Judgment.*
Philosophy Compass **10**(12), 815-827, 2015,
http://dx.doi.org/10.1111/phc3.12270,
[60] Chaplin, C.: *The Great Dictator.*
The Charles Chaplin Film Corporation, 1940,
[61] Ratcliff, C.: *The Phenomenology of Existential Feelings.*
In: Fingerhut, J. and Marienberg, S., eds.: *Feelings of Being Alive*. De Gruyter, Berlin, 2008,
[62] von Glasersfeld, E.: *Radical Constructivism: A Way of Learning.*
Routledge, London, 1995,
[63] Riegler, A.: *Towards a Radical Constructvist Understanding of Science.*
Foundations of Science **6**(1), 1-30, 2001,
http://dx.doi.org/10.1023/A:1011305022115,
[64] Schmidt, S.J.: *From Objects to Processes: A Proposal to Rewrite Radical Constructivism.*
Constructivist Foundations **7**(1), 1-9, 2011,
[65] Whitehead, A.N.: *Process and Reality: An Essay in Cosmology.*
The Free Press, New York, 1978,
[66] Froese T. and Fuchs T.: *The Extended Body: A Case Study in the Neurophenomenology of Social Interaction.*
Phenomenology and the Cognitive Science **11**(1), 205-235, 2012,
http://dx.doi.org/10.1007/s11097-012-9254-2,
[67] Lush, P., et al.: *Phenomenological control: response to imaginative suggestion predicts measures of mirror touch synaesthesia, vicarious pain and the rubber hand illusion.*
PsyArXiv, 2019,
http://dx.doi.org/10.31234/osf.io/82jav,

[68] Lah, A. and Kordeš, U.: *One Cannot "Just Ask" About Experience.*
In: Markič, O.; Strle, T.; Kordeš, U. and Gams, M., eds.: *Cognitive Sciences.* Proceedings of the 17th international multiconference "Information Society – IS 2014." Volume C. Inštitut Jožef Štefan, Ljubljana, 2014,

[69] Kordeš, U.: *Going Beyond Theory: Constructivism and Empirical Phenomenology*.
Constructivist Foundations **11**(2), 375-385, 2016,

[70] Froese, T.; Suzuki, K.; Wakisaka, S.; Ogai, Y. and Ikegami, T.: *From Artificial Life to Artificial Embodiment: Using Human-Computer Interfaces to Investigate the Embodied Mind "As-it-could-be" from the First-person Perspective.*
Proocedings of AISB **11**(1), 43-50, 2012,
http://dx.doi.org/10.1002/piuz.201290018,

[71] Hacking, I.: *The Social Construction of What?*
Harvard University Press, Cambridge, 1999,

[72] Petitmengin, C.: *Describing One's Subjective Experience in the Second Person: An Interview Method for the Science of Consciousness*.
Phenomenology and the Cognitive Sciences **5**(1), 229-269, 2006,
http://dx.doi.org/10.1007/s11097-006-9022-2,

[73] Fuchs, T.: *Embodiment and Psychopathology: A Phenomenological Perspective*.
Current Opinions in Psychiatry **22**(6), 570-575, 2009,
http://dx.doi.org/10.1097/YCO.0b013e3283318e5c,

[74] Mills, J.; Bonner, A. and Francis, K.: *The Development of Constructivist Grounded Theory.*
International Journal of Qualitative Methods **5**(1), 25-35, 2006,
http://dx.doi.org/10.1177/160940690600500103,

[75] Mills, J.; Bonner, A. and Francis, K.: *Adopting a Constructivist Approach to Grounded Theory: Implications for Research Design.*
International Journal of Nursing Practice **12**(1), 8-13, 2006,
http://dx.doi.org/10.1111/j.1440-172X.2006.00543.x.

Interdisciplinary Description of Complex Systems 19(4), 493-501, 2021

AN ARTIFICIAL IMMUNE SYSTEM APPROACH TO AUTOMATED PROGRAM VERIFICATION: TOWARDS A THEORY OF UNDECIDABILITY IN BIOLOGICAL COMPUTING

Soumya Banerjee[1, 2, *]

[1]University of Oxford
Oxford, United Kingdom

[2]Ronin Institute
Montclair, USA

DOI: 10.7906/indecs.19.4.3
Regular article

Received: 4 April 2020.
Accepted: 21 December 2021.

ABSTRACT

We propose an immune system inspired Artificial Immune System algorithm for the purposes of automated program verification. It is proposed to use this Artificial Immune System algorithm for a specific automated program verification task: that of predicting shape of program invariants. It is shown that the algorithm correctly predicts program invariant shape for a variety of benchmarked programs. Program invariants encapsulate the computability of a particular program, e.g. whether it performs a particular function correctly and whether it terminates or not. This work also lays the foundation for applying concepts of theoretical incomputability and undecidability to biological systems like the immune system that perform robust computation to eliminate pathogens.

KEY WORDS

artificial immune system, program invariant, undecidability, incomputability, biological computing, immuno-computing, fundamental limits on biological computing

CLASSIFICATION

JEL: I10, O30

*Corresponding author, η: soumya.banerjee@maths.ox.ac.uk; +1 505 277 3122;
Department of Computer Science, 1, University of New Mexico, Albuquerque, NM, 87131, USA

INTRODUCTION

The biological immune system has proved to be a rich source of inspiration for computing [1-15]. Artificial immune systems (AISs) take inspiration from the immune system to provide powerful metaphors for robust and distributed computing. In this article, I employ an immune system inspired approach to solve a problem in program verification: that of finding a program invariant.

An invariant of a program is a mathematical formula that captures the semantics of the program [16] and is used in automatic program verification. The shape of an invariant is its approximate polynomial representation. Once the shape of the invariant is predicted, deterministic techniques can be used to generate the exact form of the invariant [17]. Hence, the prediction of invariant shape is of paramount importance for program verification.

An AIS algorithmic framework is proposed to carry out the machine-learning task of predicting invariant shape from an instance of a program. Program invariants encapsulate the computability of a particular program, e.g. whether it performs a particular function correctly and whether it terminates or not. We hope this work will also lay the foundation for applying concepts of theoretical incomputability and undecidability to biological systems like the immune system that perform robust computation to eliminate pathogens [8-15].

IMMUNOLOGICAL PRELIMINARIES

A chemical species that can be recognized by the adaptive immune system is known as an antigen (*Ag*). When an organism is exposed to an *Ag*, some specialized immune system cells called B cells respond by producing chemicals called antibodies (*Ab*'s). *Ab*'s are molecules attached primarily to the surface of B cells whose aim is to recognize and bind to *Ag*'s. By binding to these *Ab*'s the Ag stimulates the B cell to proliferate and mature into plasma cells that secrete *Ab*. An organism is expected to encounter a given *Ag* repeatedly during its lifetime. The effectiveness of the immune response to secondary encounters is enhanced by the presence of memory cells associated with the first infection, capable of producing high- affinity *Ab*'s after repeat encounters. Such a strategy ensures that the speed and accuracy of the immune response becomes successively higher after each infection. This gives rise to associative memory where the stored pattern is recovered through the presentation of an incomplete version of the pattern. The repertoire of activated B cells is diversified [18-21] and B-cells with higher affinity for the antigen are selected to enter the pool of memory cells.

AUTOMATED PROGRAM VERIFICATION AND PROGRAM INVARIANTS

The field of automated program verification started with seminal work by Floyd [22] and Hoare [23]. They introduced the concept of a loop invariant: a mathematical formula that remains true throughout the execution of a loop. The loop invariant completely captures the semantics of the loop, and along with the program preconditions and postconditions, can be used to show correctness of the program [23].

Previous work [16] has shown how the loop invariant for a particular program can be generated by a priori agreement on the shape of the invariant: the approximate polynomial representation of the invariant. However, the shape of the loop invariant can be hard to deduce for many programs.

The following shows an example program:

```
{A □ 0, B □ 0}
x := A;
y := B;
z := 0;
```

```
while x > 0 do
    if odd(x) then z := z + y;
    y := 2 * y;
    x := x/2;
end while
```

Assuming the shape of the program invariant as $\mathbf{I}_{shape}$: $Ax + By + Cz + Dxy + Eyz + Fxz + Gxyz + H = 0$, (where A, B, C, D, E, F, G and H are constants or program variables), using quantifier elimination [16] the final loop invariant is Ifinal: z + xy –AB = 0. Coupled with a precondition **P**: $\{A \geq 0 \wedge B \geq 0 \wedge x = A \wedge y = B \wedge z = 0\}$ and a postcondition **Q**: $\{z = A \cdot B\}$, it can be shown that this invariant is consistent with **Q** i.e. the program correctly multiplies 2 numbers A and B and stores the result in *z*.

Finding the precise shape of the loop invariant is generally a non-trivial process and the algorithm proposed aims to use 'cues' from the program to make informed predictions about the invariant shape and ultimately help in automated program verification.

PROPOSED COMPUTATIONAL FRAMEWORK

Here we propose a computational framework for predicting program invariants. An AIS algorithm will be used to generate shapes of program invariant. Initially the AIS will be trained on programs, for which the shape of invariant is known. Then a program will be presented to the AIS and it will try to predict the form of the invariant.

An AIS approach presents many advantages over a traditional Machine Learning (ML) approach. In an AIS, recognition can be *sloppy* [24] i.e. if it has previously recognized program *P* (with an invariant *I*), then a new program *P'* 'similar' to *P*, can also be recognized, and an invariant *I'* can be generated (that is similar in form to *I*). This is akin to our immune system recognizing a previously encountered pathogen (program), and generating antibodies (invariant) similar to the previously produced antibodies.

The natural immune system produces antibodies by a process of mutation, and the same process is emulated in AIS algorithms. A candidate solution (invariant) will be generated, and then the solution will be improved by *in-silico* mutation.

Previously encountered programs and their corresponding invariants will be stored as memory B cells. When a program similar to a stored one is presented, the time taken to generate the invariant will be shorter than the time taken to generate the original invariant (*secondary response*).

COMPONENTS OF THE ARTIFICIAL IMMUNE SYSTEM

Here we define the specific components of the AIS have to be determined. What is the program analogue of an antigen and an antibody?

A *program fragment* is defined to be either an assignment statement, a statement containing an iteration construct (for, while, repeat, etc.), or a statement having a conditional check (if <condition> then) e.g. $x := x + 2$, and *while* $(x > 0)$ *do*, and *if* $(x > 3)$ *then*, are all program fragments.

The analogue of an antigen is a program fragment and the corresponding analogue of an antibody is an invariant for the program fragment it recognizes. Hence, the AIS will be presented with an antigen (program fragment), and the immune system cells will either produce the antibody (invariant) immediately if it has encountered this antigen before, or willundergo mutations to generate the correct antibody (invariant).

The individual invariants for each program fragment will then be recombined to generate the invariant for the whole program.

A SHAPE SPACE AND ANTIGENIC DISTANCE FOR PROGRAMS

We need a measure of distance between disparate program fragments, so that the AIS can recognize them and generate an antibody in response. For a natural immune system, the antibody combining region relevant to antigen binding can be specified by a number of 'shape' parameters [25] which denote the size and shape of the combining site or physical characteristics of the amino acids.

If there are *N* shape parameters, they can be combined into a vector, and antibody combining sites and antigenic determinants can be described as points **Ab** and **Ag**, in an *N* - dimensional Euclidean vector-space called *shape space* [25].

Antigenic distance between 2 antigens is the distance in shape space [26] between them e.g. $||\mathbf{Ag_1} - \mathbf{Ag_2}||$ is the distance between antigens $\mathbf{Ag_1}$ and $\mathbf{Ag_2}$ in shape space S. The *antibody distance* is the distance $||\mathbf{Ab_1} - \mathbf{Ab_2}||$ in shape space between 2 antibodies $\mathbf{Ab_1}$ and $\mathbf{Ab_2}$.

I define the *program fragment shape space* as the *N*-dimensional Euclidean vector space of program fragment characteristics like identifier name, exponent on the identifier, operator, etc. I define the corresponding *program fragment antigenic distance* as the distance $||\mathbf{P_1} - \mathbf{P_2}||$ between 2 program fragments $\mathbf{P_1}$ and $\mathbf{P_2}$ in program fragment shape space. The *program fragment antibody (invariant) distance* is the distance $||\mathbf{I_1} - \mathbf{I_2}||$ between 2 program fragments $\mathbf{I_1}$ and $\mathbf{I_2}$ in program fragment shape space.

Let us consider 2 program fragments $\mathbf{P_1}$: *x := x + 2* and $\mathbf{P_2}$: *t := t + 2*. The corresponding antibody (invariant) for $\mathbf{P_1}$ is $\mathbf{I_1}$: *x = x + 2n*, where *n* is a program variable or constant (since upon *n –1* iterations, *x* gets the value *x + 2n*). Let $\mathbf{P_1}$ and $\mathbf{I_1}$ constitute the training set. Then the AIS should be able to produce an antibody (invariant) for the program fragment $\mathbf{P_2}$ even though it has never encountered this antigen (program) before. The correct invariant is $\mathbf{I_2}$: *t = t + 2n* (where *n* is a program variable or constant) and this is indeed what the AIS generates by somatic hypermutation. The program $\mathbf{P_1}$ differs from $\mathbf{P_2}$ by 1 mutation (replacing *x* by *t* onboth sides of the assignment) i.e. the *program fragment antigenic distance* $||\mathbf{P_1} - \mathbf{P_2}||$ is 1. The invariants $\mathbf{I_1}$ and $\mathbf{I_2}$ also differ by 1 mutation (replacing *x* by *t*) i.e. the *program fragment antibody (invariant) distance* $||\mathbf{I_1} - \mathbf{I_2}||$ is 1. Hence, when an AIS trained on ($\mathbf{P_1}$, $\mathbf{I_1}$) ispresented with $\mathbf{P_2}$, it produces $\mathbf{I_2}$ using one mutation from $\mathbf{I_1}$ (Fig. 1).

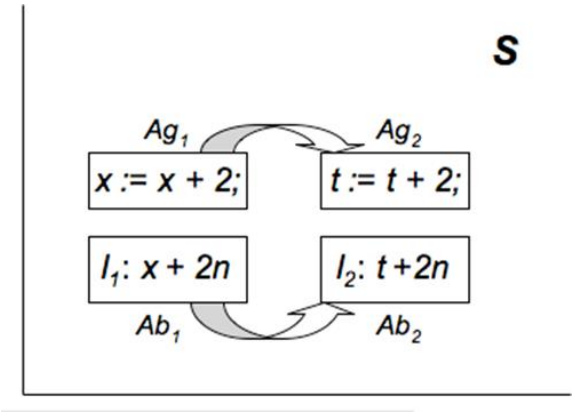

Figure. 1. AIS mutation from the assignment statement $\mathbf{Ag_1}$ (x := x + 2;) and invariant $\mathbf{Ab_1}$ (x + 2n) to $\mathbf{Ag_2}$ (t := t + 2;) and invariant $\mathbf{Ab_2}$ (t + 2n) in shape space S.

PROPOSED ALGORITHM

In this section we outline the proposed immune system inspired algorithm. The AIS would be trained on the antigen (program fragment) $\mathbf{P_1}$: *x := x + 2*and given the antibody (invariant)

$\mathbf{I_1}$: $x = x + 2n$ as a solution (*training phase*). The AIS stores the solution $\mathbf{I_1}$ as a memory detector.

When an entire program (as opposed to a program fragment) is presented to the AIS, it breaks the program up into program fragments (all the assignment statements in the program), and then 'presents' each of these antigens (fragments) to itself.

If an antigen (program fragment) $\mathbf{P_2}$ 'similar' to $\mathbf{P_1}$ is detected, it will generate $\mathbf{I_1}$ as a candidate solution. If $\mathbf{I_1}$ itself does not act as an invariant, the AIS will keep on carrying out randomly on $\mathbf{I_1}$ until it evolves the final antibody (invariant) $\mathbf{I_2}$ that will act as the invariant for the program presented (*somatic hypermutation phase*). This is akin to how the natural immune system mutates B cell receptors and ultimately produces a receptor that can recognize the antigen. The algorithm may also use some heuristics to guide the mutation process e.g. if an antigen (program fragment) of the form $p := p + 5$ is encountered, it would search its repertoire for a program fragment that is closest in program shape space to this e.g. $x := x + 5$ is closer to the presented antigen (1 mutation) than $y := y + 7$ (2 mutations). Additionally, we will have to ensure that each mutation is *sound* i.e. there is no such mutation that would generate a wrong invariant for the corresponding mutated program fragment. In the last step, the AIS incorporates $\mathbf{I_i}$ into its memory pool (*learning phase*).

The AIS then presents the next program fragment $\mathbf{P_3}$, generates the invariant $\mathbf{I_3}$ and stores it in the memory population, and so on until all program fragments have been presented. Finally, the AIS combines all invariants linearly, producing a polynomial (shape of invariant)that captures the semantics of the entire program.

RESULTS

The AIS (trained on $\mathbf{P_1}$, $\mathbf{I_1}$) presented with suites of entire programs would successfully generate the shape of the invariant. The first program is shown below:

```
(x,y,u,v) := (a,b,b,0);
x := a; y := b;
u := b; v := 0;

while (x □ y) do
    while (x > y) do x := x - y; v := v + u; end while;
    while (x < y) do y := y - x; u := u + v; end while;
end while
```

This program takes 2 positive integers *a* and *b*, and calculates their greatest common divisor and least common multiple. The AIS presents itself with each assignment statement sequentially. The first 4 assignment statements (lines 1-2) have no invariant, since they are not contained inside any loop. Hence, the AIS does not generate any invariant for them. The progress of the algorithm on the next 2 assignment statements (x := x –y; v := v + u), Fig. 2.

The AIS starts from the training set (P1: x := x + 2 & I1: x = x + 2n) and then mutates the operators and operands to create the invariant I3: x = x –yn for the program fragment P3: x := x –y. The AIS stores I3 in the memory population and for the next assignment statement (v := v + u;), it starts mutating from (P3, I3) until it creates the invariant I4: v = v + un for the program fragment P4: v := v + u.

For the next set of assignment statements (y := y –x; u := u + v;), the AIS then generates the invariants $\mathbf{I_5}$: $y = y - xn$ and $\mathbf{I_6}$: $u = u + vn$ (not shown). The 4 invariants $\mathbf{I_1}$, $\mathbf{I_2}$, $\mathbf{I_3}$ & $\mathbf{I_4}$ are then combined linearly (with *n* being substituted for all program variables, namely *x, y, u, v*) to

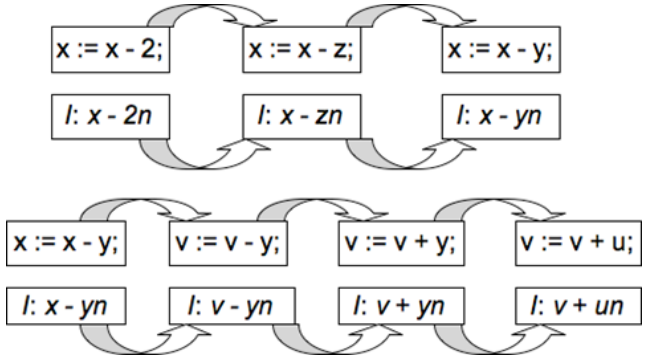

Figure. 2. AIS mutations for the assignment statements x := x –y; v := v + u.

yield the invariant shape $\mathbf{I_{shape}}$: $Ax + Bv + Cy + Du + Exy + Fy^2 + Guy + Hvy + Jxu + Ku^2 + Lvu + Mx^2 + Nvx + Pv^2 + Q = 0$, where $A, B, C, D, E, F, G, H, J, K, L, M, N, P$ and Q are constants or program variables. This is the correct invariant shape, since using quantifier elimination [16], the final invariant yielded is $\mathbf{I_{final}}$: $xu + yv - ab = 0$ (with $A = B = C = D = E = F = G = K = L = M = N = P = 0, Q = -ab, H = J = 1$).

Finally we test the AIS on another standard program [16] shown below:

```
{A ≥ 0, B ≥ 0}
x := A;
y := B;
z := 1;

while y > 0 do
    if odd(y) then y := y - 1; z := x * z;
    else x := x * x;    y := y/2;
end while
```

This program calculates A^B and stores it in z. The AIS would calculate the invariant for the program fragment $\mathbf{P_5}$: $z := x * z$ as $\mathbf{I_5}$: $z = x^n * z$. For the program fragment $\mathbf{P_6}$: x := x * x, it generates the invariant $\mathbf{I_6}$: $x = exp(x, exp(2,n))$, where $exp()$ is the exponentiation function. Combining all the program fragment invariants, gives us the following invariant shape:
$\mathbf{I_{shape}}$: $Azx^x + Bzx^y + Czx^z + D.exp(x,exp(2,x)) + E.exp(x,exp(2,y)) + F.exp(x,exp(2,z)) + G = 0$.

This is the exact shape of the invariants, since quantifier elimination yields the final invariant $\mathbf{I_{final}}$: $zx^y = A^B$ (with $A = C = D = E = F = 0, G = -A^B$).

We can now readily verify the working of the program. When the loop terminates, the invariant is true and $y = 0$, which yields the correct postcondition: $z = A^B$.

The proposed algorithm would use a sequence of mutations, guided by heuristics, to generate the correct invariant for a program invariant.

CONCLUSION AND FUTURE WORK

We have proposed a computational framework for an immune system inspired approach for automated program verification. The immune system inspired algorithm breaks up a program into fragments and presents them to itself. It then generates an invariant in response to each program fragment and ultimately combines them to create the general shape of the invariant. We show how this approach can be used to generate the general form of the program invariant for non-trivial benchmark programs [16].

Future work will focus on theoretical research into whether there are classes of programs

for which a linear combination of individual program fragment invariants might not generate the invariant for the entire program. Another avenue of future investigation would be to look into how mutations on exponentiation would affect the invariant e.g. $x := x + 2$ getting mutated to $x := x^2 + 2$. Lastly, our approach does not consider program fragments having iteration constructs like *while*, *repeat*, etc. and future research will investigate how incorporation of such program fragments can enhance the predictive power of the algorithm.

A lot of work has been done on incomputability, undecidability and program termination in theoretical computer science. The best characterization of this comes in the form of the Halting Problem formulated by Alan Turing. Biological systems also perform computing, e.g. the immune system computes the most efficient way to eliminate pathogens in a timely manner without harming the host [8-15]. However it has been more difficult to define incomputability and undecidability for biological systems.

Program invariants encapsulate the computability and correctness of a particular program, e.g. what it does and whether it terminates or not. This work lays the foundation of applying computability to biological systems especially the immune system that performs computation.

The present work also applies immune system inspired algorithms to find program invariants and prove correctness and termination. It is intriguing to speculate that it is also possible to go in the reverse direction and translate the complexities of the immune system into an equivalent computer program. The translated computer program can then be analyzed for mathematical properties of what it computes [27, 28]. Hence this work can be extended to provide a theoretical framework for understanding the limits of computation in the immune system.

The present computational framework can be used to account for cases when the immune system fails to clear infections as is the case in certain virulent infections [29].

This approach can also be similarly extended to analyse substrates for computing that are non-silicon based and can be used to probe the computational nature of life itself [15].

In summary, the present work applies the theoretical concepts of undecidability to immuno- computing and possibly biological computing in general. We view this work as the first step towards elucidating the fundamental limits of computing in immunology and possibly biology as well.

ACKNOWLEDGEMENTS

The author wishes to thank Dr. Sara Jane-Dunn, Dr. Boyan Yordanov, Prof. Deepak Kapur and Dr. ThanhVu Nguyen for helpful comments.

REFERENCES

[1] de Castro, L.N. and Timmis, J.: *Artificial Immune Systems: A New Computational Intelligence Approach.*
Springer-Verlag, London, 1996,

[2] Hunt, J.E. and Cooke, D.E.: *Learning using an artificial immune system.*
Journal of Network and Computer Applications **19**(2), 189-212, 1996,
http://dx.doi.org/10.1006/jnca.1996.0014,

[3] Dasgupta, D.: *Artificial Immune Systems and Their Applications.*
Springer-Verlag, Berlin, 1999,

[4] Hofmeyr, S.A. and Forrest, S.: *Immunity by design: An artificial immune system.*
Proceedings Genetic and Evolutionary Computation conference, pp.1289-1296, 1999,

[5] de Castro, L.N. and Von Zuben, F.J.: *Artificial Immune Systems: Part I. Basic Theory and Applications.*
School of Electrical and Computer Engineering, University of Campinas, Campinas, 1999,

[6] de Castro, L.N. and Von Zuben, F.J.: *Artificial Immune Systems: Part II – A Survey of Applications.*
School of Electrical and Computer Engineering, University of Campinas, Campinas, 2000,
[7] de Castro, L.N. and von Zuben, F.J.: *Learning and Optimization Using the Clonal Selection Principle.*
IEEE Transactions on Evolutionary Computation **6**(3), 239-251, 2002,
http://dx.doi.org/10.1109/TEVC.2002.1011539,
[8] Banerjee, S. and Moses, M.: *Scale Invariance of Immune System Response Rates and Times: Perspectives on Immune System Architecture and Implications for Artificial Immune Systems.*
Swarm Intelligence **4**, 301-318, 2010,
http://dx.doi.org/10.1007/s11721-010-0048-2,
[9] Banerjee, S.: *Scaling in the immune system.* Ph.D. Thesis.
University of New Mexico, Albuquerque, 2013,
[10] Banerjee, S., et. al.: *The Value of Inflammatory Signals in Adaptive Immune Responses.*
The 10th International Conference on Artificial Immune Systems, 2011,
[11] Drew, L., et al.: *A spatial model of the efficiency of T cell search in the influenza-infected lung.*
Journal of Theoretical Biology **398**(7), 52-63, 2016,
http://dx.doi.org/10.1016/j.jtbi.2016.02.022,
[12] Banerjee, S.: *A Biologically Inspired Model of Distributed Online Communication Supporting Efficient Search and Diffusion of Innovation.*
Interdisciplinary Description of Complex Systems **14**(1), 10-22, 2016,
http://dx.doi.org/10.7906/indecs.14.1.2,
[13] Banerjee, S.; van Hentenryck, P. and Cebrian, M.: *Competitive dynamics between criminals and law enforcement explains the super-linear scaling of crime in cities.*
Palgrave Communications **1**, No. 15022, 2015,
http://dx.doi.org/10.1057/palcomms.2015.22,
[14] Banerjee, S. and Hecker, J.P.: *A Multi-Agent System Approach to Load-Balancing and Resource Allocation for Distributed Computing.*
Complex Systems Digital Campus, World e-Conference, Conference on Complex Systems, 2015,
[15] Banerjee, S.: *A Roadmap for a Computational Theory of the Value of Information in Origin of Life Questions.*
Interdisciplinary Description of Complex Systems **14**(3), 314-321, 2016,
http://dx.doi.org/10.7906/indecs.14.3.4,
[16] Kapur, D.: *Automatically Generating Loop Invariants Using Quantifier Elimination.*
IMACS International IMACS Conference on Applications of Computer Algebra, 2004,
[17] Rodriguez, E. and Kapur, D.: *Automatic Generation of Polynomial Loop Invariants: Algebraic Foundations.*
International Conference on Symbolic and Algebraic Computation, 2004,
[18] Berek, C. and Ziegner, M.: *The maturation of the immune response.*
Immunology Today **14**(8), 400-404, 1993,
http://dx.doi.org/10.1016/0167-5699(93)90143-9,
[19] George, A.J.T. and Gray, D.: *Receptor editing during affinity maturation.*
Immunology Today **20**(4), 196, 1999,
http://dx.doi.org/10.1016/S0167-5699(98)01408-X,
[20] Nussenzweig, M.C.: *Immune receptor editing: Revise and select.*
Cell **95**(7), 875-878, 1998,
http://dx.doi.org/10.1016/S0092-8674(00)81711-0,
[21] Tonegawa, S.: *Somatic generation of antibody diversity.*
Nature **302**(14), 575-581, 1983,
http://dx.doi.org/10.1038/302575a0,

[22] Floyd, R.W.: *Assigning meanings to programs.*
Proceedings of the American Mathematical Society Symposia on Applied Mathematics **19**, 19-31, 1967,

[23] Hoare, C.A.R.: *An axiomatic basis for computer programming.*
Communications of the ACM **12**(10), 576-585, 1969,
http://dx.doi.org/10.1145/363235.363259,

[24] Perelson, A.S. and Wiegel, F.W.: *Some design principles for immune system recognition.*
Complexity, 1999,

[25] Perelson, A.S. and Oster, G.F.: *Theoretical Studies of Clonal Selection: Minimal Antibody Repertoire Size and Reliability of Self-Non-self Discrimination.*
Journal of Theoretical Biology **81**(4), 645-67, 1979,

[26] Smith, D.J.; Forrest, S.; Hightower, R.R. and Perelson, S.A.: *Deriving shape space parameters from immunological data.*
Journal of Theoretical Biology **189**(2), 141-150, 1997,
http://dx.doi.org/10.1006/jtbi.1997.0495,

[27] Dunn, S.J., et al.: *Defining an essential transcription factor program for naïve pluripotency.*
Science **344**(6188), 1156-1160, 2014,
http://dx.doi.org/10.1126/science.1248882,

[28] Yordanov, B., et al.: *A method to identify and analyze biological programs through automated reasoning.*
npj Systems Biology and Applications **2**, No. 16010, 2016,
http://dx.doi.org/10.1038/npjsba.2016.10,

[29] Banerjee, S. et al.: *Estimating Biologically Relevant Parameters under Uncertainty for Experimental Within-Host Murine West Nile Virus Infection.*
Journal of the Royal Society Interface **13**(117), No. 20160130, 2016,
http://dx.doi.org/10.1098/rsif.2016.0130.

Interdisciplinary Description of Complex Systems 19(4), 502-510, 2021

LYMPH NODE INSPIRED COMPUTING: TOWARDS IMMUNE SYSTEM INSPIRED HUMAN-ENGINEERED COMPLEX SYSTEMS

Soumya Banerjee*

University of Cambridge
Cambridge, United Kingdom

DOI: 10.7906/indecs.19.4.4
Regular article

Received: 26 September 2021.
Accepted: 21 December 2021.

ABSTRACT

The immune system is a distributed decentralized system that functions without any centralized control. The immune system has millions of cells that function somewhat independently and can detect and respond to pathogens with considerable speed and efficiency. Lymph nodes are physical anatomical structures that allow the immune system to rapidly detect pathogens and mobilize cells to respond to it. Lymph nodes function as: 1) information processing centres, and 2) a distributed detection and response network. We introduce biologically inspired computing that uses lymph nodes as inspiration. We outline applications to diverse domains like mobile robots, distributed computing clusters, peer-to-peer networks and online social networks. We argue that lymph node inspired computing systems provide powerful metaphors for distributed computing and complement existing artificial immune systems. We view our work as a first step towards holistic simulations of the immune system that would capture all the complexities and the power of a complex adaptive system like the immune system. Ultimately this would lead to immune system inspired computing that captures all the complexities and power of the immune system in human-engineered complex systems.

KEY WORDS

biologically inspired computing, artificial immune systems, computational immunology, lymph node computing

CLASSIFICATION

JEL: I12, O30

*Corresponding author, η: soumya.banerjee@maths.ox.ac.uk; +1 505 277 3122;
Department of Computer Science, 1, University of New Mexico, Albuquerque, NM, 87131, USA

INTRODUCTION

The immune system is a distributed decentralized system that functions without any centralized control. The immune system has millions of cells that function somewhat independently and can detect and respond to pathogens with considerable speed and efficiency [1-3]. Lymph nodes are physical anatomical structures that allow the immune system to rapidly detect pathogens and mobilize cells to respond to it. Lymph nodes function as: 1) Information processing centres, and 2) Distributed detection and response network.

We introduce biologically inspired computing that uses lymph nodes as inspiration. Lymph nodes process information. They facilitate detection of pathogens, process that information and help co-ordinate the output or response to infections. The immune system is a complex adaptive system that has evolved for millions of years under co-evolutionary pressure from pathogens. It has evolved mechanisms to efficiently detect and respond to pathogens.

Innovations include lymph nodes which serve as information processing centres: they integrate signals from neighbouring tissues, process it, match pathogens with immune system cells and then send out cells to respond to the pathogen. A schematic of this is shown in Fig. 1.

Only a few cells (1 in a million) of the immune system can recognize a specific pathogen [1]. Once a particular immune system cell detects a pathogen, it must replicate and build copies of itself and then act against the pathogen. Additionally, detection and the response to pathogens must be rapid otherwise the pathogen will be able to replicate and kill the host organism.

Lymph nodes serve to facilitate the serendipitous encounter of rare immune system cells with their antigen.

Response against the pathogen is complicated by two considerations of body size of the animal (the size of the system):

1. Certain cells of the immune system release chemicals called antibodies which are further diluted in blood. Hence in larger animals, antibodies get diluted more in the larger volume of blood. Hence larger animals must secrete more absolute quantities of antibodies [1].
2. Certain cells of the immune system physically search for infected cells. This search is more difficult in larger animals since the search is through a larger physical space [4-6].

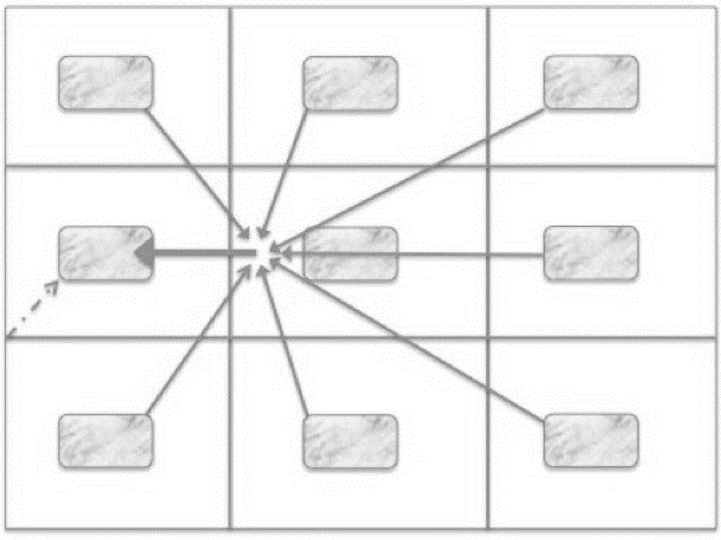

Figure 1. A schematic of how a network of lymph nodes co-ordinates and process information locally and globally to respond to infections.

We can draw valuable inspiration from how the immune system has designed a distributed detection and response infrastructure. The *hardware* of the immune system (lymph nodes) [1] has co-evolved with its *software* (T-cell recirculation and search strategies) [4, 6] to enable the immune system to mount effective and timely responses to pathogens. Here we outline applications to diverse human-engineered complex systems like social networks, intrusion detection in distributed systems, cluster allocation in distributed systems of computers and mobile robots.

We argue that the biological immune system can provide powerful metaphors and inspiration for distributed computing in human-engineered complex systems like clusters of computers, peer-to-peer networks and online social networks. Lymph node inspired techniques can also complement and provide a powerful context in which to situate other immune system inspired techniques like negative selection [7], clonal selection [8] and dendritic cell algorithms [9].
In the following sections, we outline some biological preliminaries and the general principles of lymph node inspired computing. We lay out applications to human-engineered systems, fault-tolerant distributed computing, computer security and other resource constrained distributed systems.

BIOLOGICAL PRELIMINARIES

The draining region of a lymph node is the region of tissue adjoining a lymph node. Dendritic cells patrolling the draining region home to the nearest lymph node to present antigen. Cells of the immune system called dendritic cells search for pathogens in tissue and once they find it, home to lymph nodes. Within lymph nodes they present antigen to other specialized immune system cells called T-cells and B-cells. Almost 1 in a million T-cells and B-cells are able to recognize this antigen [1]. Once the antigen is recognized, T-cells exit the lymph node and home to tissues to find infected cells. B-cells go through a similar process and secrete chemicals called antibodies that neutralize pathogens.

Previous work has shown that there needs to be a balance between the local time of detection of pathogen with global time for response against the infection (by secreting antibodies) [1-3]. The optimal architecture of lymph nodes needs to be sub-modular in order to ensure that the time to detect and respond does not scale appreciably with the size of the system (in this case the body size of the animal). Each lymph node has a *protection* (with one copy of each immune system cell specific to pathogens) [1]. This is a modular search unit which is iterated in larger animals.

If cognate immune system cells are so rare, then how does the immune system detect pathogens quickly and mount an effective immune response? The answer lies in lymph nodes that are:

1. modular,
2. parallel, and
3. privileged metabolism: The metabolic rate of lymph nodes is invariant with the size of the host animal [10].

APPLICATIONS TO FAULT-TOLERANT DISTRIBUTED COMPUTING AND SEARCH IN SOCIAL NETWORKS

MODULAR SEARCH IN DISTRIBUTED SYSTEMS OF COMPUTERS

Lymph node inspired computing can be applied to a system where controllers try to find computer clusters that can be used for computation. Let an artificial lymph node be composed of a number of clusters and a process queue that will manage requests to schedule programs

on a cluster of computers [11]. Also let there be a number of such artificial lymph nodes that have the capability of communicating with each other. An artificial lymph node is a computer in charge of a number of clusters. This computer will store the process queue and also will have some memory and computing power to communicate with other artificial lymph nodes.

We seek to minimize the total time to find a cluster. There is a trade-off between the local cost of traversing through the queue in a lymph node which is $O(n^2)$ and the global cost of communicating with other lymph nodes which is $O(N/n)$. Here n is the number of clusters in a single lymph node and N is the total number of artificial lymph nodes in the complete system.

We assume that the global cost of finding another cluster in another lymph node that can service some process requirement is proportional to the number of artificial lymph nodes (where N/n is the number of artificial lymph nodes in the system). Minimizing the total time cost, we get:

$$n = O(N^{1/3}).$$

This implies that in larger systems, the number of clusters within a single lymph node should grow larger but only sub-linearly in the number of total clusters in the system. Hence a lymph node inspired approach would balance local costs of queue traversal and global costs of finding artificial lymph nodes with another cluster that can service the process [11].

IMMUNE SYSTEM INSPIRED DECENTRALIZED SEARCH IN SOCIAL NETWORKS

Social networks are characterized by short diameters and it is very easy to find another person within about 5 to 6 hops [12]. This is exemplified by experiments performed by Stanley Milgram where letters posted by complete strangers found their way to people across vast geographical distances [12]. Kleinberg [13] gave a very elegant explanation of this phenomenon based on the small-world degree connection of nodes in a network.

The work implicitly ignored considerations of space, since long-distance links are assumed to have the same cost as short-distance links. Work in the immune system is motivated primarily by constraints of space. An infected site lymph node has to incur increasing communication costs in larger organisms, since it has to recruit immune system cells over larger physical distances (and hence long-distance links have higher associated costs).

Real world social networks also densify i.e. the average number of neighbours that an individual has increases with time. Hence such networks will have a communication cost not only due to space but due to the requirement of maintaining a certain number of connections. In work inspired by conceptual similarities between a lymph node and a social community, a non-spatial communication cost was incorporated in order to introduce the realism of individuals communicating within communities [14].

It has been shown that the optimal strategy that minimizes the time to find someone using only local information is where the size of communities and the number of communities both increase with the size of the system. This is ultimately predicted to lead a speedup in search for information. Such a strategy would decrease the time required to search for rare information in online social networks [14]. Figure 2 shows example of a lymph node inspired social network architecture.

APPLICATIONS TO COMPUTER SECURITY AND RESOURCE-CONSTRAINED DISTRIBUTED SYSTEMS

INTRUSION DETECTION SYSTEMS

Lymph node inspired strategies can be used to augment intrusion detection applications like LISYS [15] and process Homeostasis (pH) [16]. pH adaptively reduces computer processing

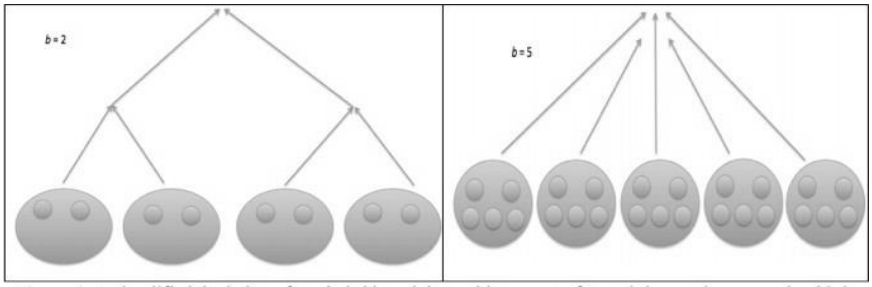

Figure 2. A lymph node inspired social network architecture that is optimal for detecting information (adapted from [14]). Left Panel: social network with 2 persons per community (similar to an artificial lymph node). Right Panel: a scaled up social network with 5 persons per community (artificial lymph node grows in size as the network gets larger).

speed to reduce the spread of malware. A lymph node inspired approach can be used to augment pH in the following manner: a subnet of computers would be an artificial draining region and a security node in charge of the subnet would be analogous to a lymph node.

A lymph node inspired architecture can balance local communication (intrusion detection) and global communication (alert propagation). We note that in some cases it may not be advisable to spread alerts immediately in pH [16], since the performance of the whole system may degrade or false alerts can propagate rapidly throughout the system.

LOW POWER RESOURCE CONSTRAINED DISTRIBUTED SYSTEMS

A lymph node inspired approach may be particularly useful for security of low power distributed systems like mobile phones. Mobile phone viruses propagate by small hops to neighbouring within-range devices, and hence physical space and proximity of devices is important [17]. Since mobile phones are constrained to communicate to nearby neighbours (through Bluetooth connections and proximal cell towers), the lymph node analogy can be extended to view mobile phone transmission towers as artificial lymph nodes and the area of mobile phone users serviced by it as the draining region.

APPLICATION TO HUMAN-ENGINEERED SYSTEMS

MODULAR SEARCH IN MULTI-ROBOT SYSTEMS

Our work can also be applied to an immune system inspired robotics systems. The draining region of a lymph node can be patrolled by robots. The artificial lymph node is a central computer. Robots in the draining region communicate with the central computer (artificial lymph node). A lymph node inspired architecture can lead to speedup in search times for robots [1-4] by balancing the local time to find a solution (similar to a pathogen) and the global time to propagate it to other artificial lymph nodes.

SEARCH IN PEER-TO-PEER SYSTEMS

Distributed peer-to-peer systems are used to provide services like search and content integration. Computer nodes store data or service and no single node has complete global information. Decentralized search using local information is used to locate data. In this case, artificial lymph nodes would be clusters of computer nodes that store information. Artificial lymph nodes communicate with each other when they need to find information. Using a lymph node inspired can lead to speedup in time to search for information in peer-to-peer systems [3].

GENERAL PRINCIPLES OF LYMPH NODE INSPIRED COMPUTING

Here we outline the general principles of lymph node inspired computing. The immune system aims to balance the local time of detection of pathogens with the global time for response against pathogens.

BALANCE BETWEEN LOCAL DETECTION AND GLOBAL RESPONSE

The key argument is that response against pathogens by the immune system should minimize energy expenditure. This leads to a particular scaling of the size and number of lymph nodes: a relationship for how the size and number of lymph nodes should vary with the size of the animal [1]. We assume that energy spent is linear in the size of the organism, i.e. all organisms spend a constant fraction of their energy budget on the immune system. Similarly we strive to build systems that have approximately linear or poly-logarithmic scaling in terms of memory or computation requirements.

MISCELLANEOUS CONSIDERATIONS

Any distributed computer system with a client-server relationship will lead to different kinds of tradeoffs between local and global communication. Such systems could have a master-slave relationship in which individual components report aggregated data to a processing centre (local communication) and processing centres distribute data globally among all components (global communication) [1-3]. A lymph node inspired strategy can enhance message propagation times and increase robustness in such systems. The final optimal architecture will also depend on architecture of the system and the idiosyncrasies of the system; for example in certain intrusion detection systems like process Homeostasis [16] it may not be optimal to spread information globally (which would lead to performance degradation of the whole system) and hence the optimal strategy would be to contain all information locally.

GENERAL THEORY OF LYMPH NODE INSPIRED SEARCH AND RESPONSE IN DISTRIBUTED SYSTEMS

Generally, if the local and global communication costs scale with exponents α and β, we have the time to find information locally and then spread it globally as

$$t_{\text{total}} = \mathrm{O}(n^{\alpha}) + \mathrm{O}(N^{\gamma} / n^{\beta}),$$

where n is the number of computers in an artificial lymph node and N is the number of artificial lymph nodes. Minimizing the expression with respect to N, and assuming a power-law scaling relation, we get the following general relation [11] for the number of nodes or computers within an artificial lymph node:

$$n = \mathrm{O}(N^{\gamma/(\alpha+\beta)})$$

We also have the following regimes based on the relative values of the parameters:

1. if $\gamma < \alpha + \beta$ we have sub-linear scaling,
2. if $\gamma > \alpha + \beta$ we have super-linear scaling,
3. if $\gamma = \alpha + \beta$ we have linear scaling,
4. if $\gamma/(\alpha + \beta) = 0$ we have no scaling (constant),
5. if $\gamma/(\alpha + \beta) < 0$ we have negative scaling.

DISCUSSION

Lymph nodes are physical anatomical structures that allows the immune system to rapidly detect pathogens and mobilize cells to respond to it. Lymph nodes function as:

1. information processing centres, and
2. distributed detection and response networks.

We introduce biologically inspired computing that uses lymph nodes as inspiration. This is shown schematically in Fig. 3.

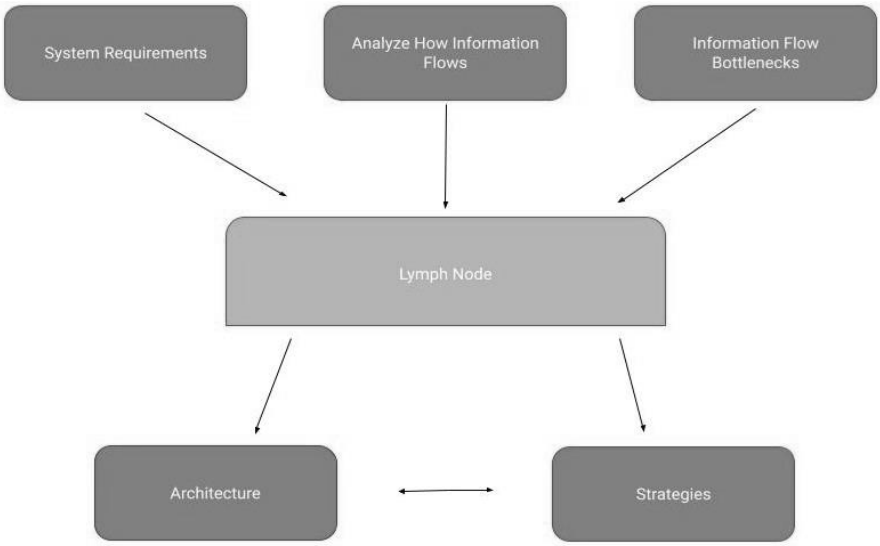

Figure 3. A lymph node produces a unique informational lens through which to study immune systems and also distributed systems. One can study the requirements of a system that needs to be built, analyse its information flow and then propose a lymph node inspired architecture and strategy to solve that problem.

The immune system is able to and rare spatially localized pathogens and eliminate them in a timely manner. The immune system uses specialized cells to find pathogens in anatomical regions called lymph nodes. A sub-modular arrangement of lymph nodes could lead to fast elimination of pathogens in the immune system and also faster search for solutions in immune inspired distributed systems of computers [1-5]. Each lymph node has a modular unit of protection called a *protection* (that contains the 1 in a million immune system cells specific to a pathogen). This is a modular search unit that may be iterated in larger animals and allows the immune system to detect pathogens quickly and mount an effective immune response.

Our approach highlights the tight coupling and co-evolution of hardware (lymph nodes) vs. software (algorithms for guiding T-cells to sites of infection) in complex biological systems like the immune system. The software would be specific algorithms or search processes that guide T-cells to infected cells in tissue. The hardware is the physical infrastructure of lymph nodes and circulatory networks. Both have co-evolved with each other to produce a system that is capable of detecting an event and responding to it efficiently.

Human-engineered complex systems can draw valuable inspiration from such systems. The *hardware* of the immune system (lymph nodes) [1] has co-evolved with its *software* (T-cell recirculation and search strategies) to enable the immune system to mount effective and timely responses to pathogens. Here we have shown how these techniques can be applied to systems like mobile networks and outline applications to diverse human-engineered complex systems like social networks, intrusion detection in distributed systems, cluster allocation in distributed systems of computers and mobile robots.

We argue that the biological immune system can provide powerful metaphors and inspiration for distributed computing in human-engineered complex systems like clusters of computers, peer-to-peer networks and online social networks. Such a lymph node inspired approach can also be applied to the Internet of Things.

Lymph node inspired techniques can also complement and provide a powerful context in which to situate other immune system inspired techniques like negative selection [7], clonal selection [8] and dendritic cell algorithms [9]. These techniques can provide valuable inspiration for building systems that scale gracefully as we increase the size of the system.

We hope that this would lead to immune system inspired computing that captures all the complexities and power of the immune system in human-engineered complex systems.

REFERENCES

[1] Banerjee, S. and Moses, M.: *Scale: Invariance of Immune System Response Rates and Times: Perspectives on Immune System Architecture and Implications for Artificial Immune Systems*.
Swarm Intelligence, 2010,

[2] Banerjee, S. and Moses, M.: *A hybrid agent based and differential equation model of body size effects on pathogen replication and immune system response.*
In: Andrews, P.S., et al., eds.: *Artificial immune systems*. Lecture notes in computer science **5666**. Springer, Berlin, pp.14-18, 2009,

[3] Banerjee, S. and Moses, M.: *Modular RADAR: An immune system inspired search and response strategy for distributed systems.*
In: Hart, E., et al., eds.: *Artificial immune systems*. Lecture notes in computer science **6209**. Springer, Berlin, pp.116-129, 2010,

[4] Banerjee, S.; Levin, D.; Moses, M.; Koster, F. and Forrest, S.: *The Value of Inflammatory Signals in Adaptive Immune Responses.*
10th International Conference on Artificial Immune Systems (ICARIS), 2011,

[5] Banerjee, S.: *Scaling in the Immune System.* Ph.D. Thesis.
University of New Mexico, New Mexico, 2013,

[6] Levin, D., et al.: *A spatial model of the efficiency of T cell search in the influenza-infected lung.*
Journal of Theoretical Biology **398**, 52-63, 2016,
http://dx.doi.org/10.1016/j.jtbi.2016.02.022,

[7] Forrest, S.; Perelson, A.; Allen, L. and Cherukuri, R.: *Self-nonself discrimination in a computer*.
In: *Proceedings of the 1994 IEEE Symposium on Research in Security and Privacy*. IEEE Computer Society Press, Los Alamitos, pp.202-212, 1994,

[8] De Castro, L.N. and Von Zuben, F.J.: *Learning and optimization using the clonal selection principle*.
IEEE Transactions on Evolutionary Computation **6**(3), 239-251, 2002,
http://dx.doi.org/10.1109/TEVC.2002.1011539,

[9] Greensmith, J.; Aickelin, U. and Cayzer, S.: *Introducing dendritic cells as a novel immune-inspired algorithm for anomaly detection.*
In: *International Conference on Artificial Immune Systems.* Springer, Berlin, pp.153-167, 2005,

[10] Wiegel, F.W. and Perelson, A.: *Some scaling principles for the immune system.*
Immunology and Cell Biology **82**(2), 127-131, 2004,
http://dx.doi.org/10.1046/j.0818-9641.2004.01229.x,

[11] Banerjee, S. and Hecker, J.: *A Multi-Agent System Approach to Load-Balancing and Resource Allocation for Distributed Computing*.
arXiv preprint:1509.06420, 2015,

[12] Milgram, S.: *The small world problem*.
Psychology Today **2**, 60-67, 1967,

[13] Kleinberg, J.M.: *The Small-World Phenomenon: An Algorithmic Perspective.*
In: Yao, F. and Luks, E., eds.: *Proceedings of the thirty-second annual ACM symposium on Theory of computing*. ACM, New York, 2000,

[14] Banerjee, S.: *A Biologically Inspired Model of Distributed Online Communication Supporting Efficient Search and Diffusion of Innovation.*
Interdisciplinary Description of Complex Systems **14**(1), 10-22, 2016,
http://dx.doi.org/10.7906/indecs.14.1.2,
[15] Somayaji, A. and Forrest, S.: *Automated response using system-call delays.*
Usenix Security Symposium, 2000,
[16] Hofmeyr, S.A. and Forrest, S.: *Architecture for an artificial immune system.*
Evolutionary Computation **8**(4), 443-473, 2000,
http://dx.doi.org/10.1162/106365600568257,
[17] Kleinberg, J.: *Computing: The wireless epidemic.*
Nature **449**, 287-288, 2007,
http://dx.doi.org/10.1038/449287a.

Interdisciplinary Description of Complex Systems 19(4), 511-525, 2021

CONTENT ANALYSIS OF JOB ADVERTISEMENTS FOR IDENTIFYING EMPLOYABILITY SKILLS

Ivona Lipovac and Marina Bagić Babac*

University of Zagreb, Faculty of Electrical Engineering and Computing
Zagreb, Croatia

DOI: 10.7906/indecs.19.4.5
Regular article

Received: 16 August 2021.
Accepted: 26 November 2021.

ABSTRACT

Recently, web-based job sites appear to be the major source of advertisements. However, it has been argued that search engine toolkits restrict wording used to find job skill requirements, which requires clearly defined commonly used terms. This article aims to analyse 16 000 online job advertisements using content analysis to identify current skills required by various professions in order to allow the comparison across countries through time to identify if trends vary in different national markets. In addition, algorithmic approaches to processing data were used to collect, sift, and organize content as well as to improve the efficiency and reliability of the analysis, and to derive a conceptual model of practitioner knowledge, skills, and abilities. Useful insights were given about the skills that employers require, based on online job advertisements from the USA, UK, Ireland, and Hong Kong.

KEY WORDS

content analysis, job advertisements, workforce skills

CLASSIFICATION

JEL: J23, J24

*Corresponding author, η: marina.bagic@fer.hr; +385 1 6129864;
FER, Unska 3, HR-10000 Zagreb

INTRODUCTION

The labour market is extensive and exposed to constant changes due to globalization, development of technology and science, changes in the demographic structure and environment. It is important to know labour market, its trends, employers' requirements, for an individual to easier and better prepare for the desired job.

The research which uses job advertisements as data are growing in popularity in the new era. This type of research provides useful insight into the labour market and gives an opportunity to understand working conditions, salaries, and occupational changes. A major motivation for these kinds of studies is to examine the changing nature of skills that are required in the workplace. A development like this could enable more collaboration between researchers and recruiting organisations resulting in major benefits for both recruiters and individuals seeking for the job [1].

In the study conducted by Maier et al. [2], the researchers discovered that the number and variety of technical skills mentioned in job ads were increasing over the time examined. This indicates that the desired skills of employees are evolving over time, so knowing employers' requirements helps individuals to remain competent. In addition to the facts stated above, knowing labour market, and required skills are not only important for individuals seeking for the job, but also for policymakers in the areas of education and training, career guidance, labour market activation, immigration, and enterprise development, employers, HR consultants and recruiters [3].

The primary goal of this article is to provide trends into employment and workforce skills as such indicators can help to influence rising educational attainment, technologic convergence, and demographic change. In addition, these trends indicate skill shortages, consumer demand, and changing industrial structure. Therefore, a content analysis of job advertisements is used to identify current skills required by various professions in order to allow the comparison across countries through time to identify if trends vary in different national markets.

This article is organized into four major sections. Related research is reviewed in the introductory section. The following section describes the methods and necessary preparations made before the analysis itself. Finally, the results achieved through the content analysis are presented and discussed, with several conclusions drawn.

RELATED WORK

A primary goal motivating the content analysis of job advertisements in numerous research studies has been to derive a conceptual model of employees' knowledge, skills, and abilities [3]. From the study by Maier and Clark [2], it is noticeable that this kind of research is not so new, indicating how job advertisements were analysed across different decades.

Using content analysis of online job advertisements from *Monster.com*, Backhaus [4] argues that differences exist among companies in recruitment tactics. For instance, certain companies focus on company branding rather than on employee benefits. Baravalle and Capiluppi [5] used the same website in their research as Backhaus. They focused on job advertisements in IT sector and discovered a very common problem of mismatch between requirements of UK industry and offer of educational and training institutions.

To derive trends in programming skills, Smith, and Ali [6] examined job advertisements from online job agency *www.dice.com*. Their approach included extracting key terms from over 80 000 job advertisements. The *Dice* platform was also used in research by Surakka [7] with similar motivation for research as Smith and Ali [6], reporting that the duties of software developers changed as technically more versatile, that it is no longer enough to have skills only in one or

two programming languages. Molinero and Xie [8] presented key groups of skills in online job advertisements using cluster analysis in combination with multi-dimensional scaling.

Lee and Lee [9] gathered job advertisements from websites of companies featured in Fortune 500. The focus of their research was on job advertisements for IT managers. The outcomes of the research indicate that IT job applicants with high school diplomas or associate degrees possibly would not become IT managers in these companies because these companies do not appreciate the value of certification but strongly require their IT managers to possess technical skills and system skills, as well as business skills.

Furthermore, Dörfler and Werfhorst [10] focused on skills that employers require in various occupations. The outcome of their research is that employers require a wider set of skills over time, including not only occupational skills but also social and personal skills. A similar conclusion was presented in research by Kureková, Beblavý, and Thum-Thysen [11]. Using content analysis and simple statistical methods, they explored the Slovak labour market and found that employers in Slovakia are fairly demanding a wide set of skills even in formally low-skilled jobs. Using text mining on publicly available job advertisements, Pejić Bach [12] performed research related to analysis of competencies required in Industry 4.0 to develop a profile of Industry 4.0 job advertisements. However, Bennett [13] argues that the problem with skills employers' demand is that usually the level of competence for each skill is not defined, which leads to problems where universities do not know what to teach and candidates do not know exactly what they are being asked to demonstrate. Thus, Bennett suggests that organisations create a uniform set of short, straightforward, and easily memorable definitions of key skills to facilitate shared understanding.

Several research studies focused on exploring the quality of online job advertisements as a data source. Kureková, Beblavý, and Thum-Thysen [11] suggested strategies for overcoming selected methodological issues and that online job advertisements can be coupled with other sources of vacancy data or text describing analysed professions. One of the goals of the research by Wade and Parent [14] was to determine how the required skill mix and the degree to which subjective assessments of the possession of skills affect the job performance. Also, researching online job advertisements to identify the mix of desired skills is very useful and can be exploited as valuable input to student counselling services or curriculum development. In addition, Huang [15] points out that the online job advertisements list a wider mix of skills, while practitioner literature tends to focus on technical skills.

There are numerous research studies conducted as a tool for improving education, training, and performance management such as a study by Iyer [16] used a content analysis of 394 job announcements in the visual resources field to improve education and training by providing the library community with the information necessary to support the development of programs for visual resource professionals. Another research with a similar purpose was the one by Payne [17]. The study examined Information Management (IM) within the UK through the comparison of content analyses of IM course curricula and IM job advertisements to determine what is recognised as the discipline of IM through the consideration of what components are taught within UK Higher Education Information Management degree curricula. Reeves and Hahn [18] also highlight the fact that the content analysis of job advertisements can inform curriculum development and enhancement, academic advising, and job-seeking implications for new graduates. They also suggest that these kinds of studies are needed all the time to reflect the current state of the labour market. The results of the previously mentioned research by Surakka [7] indicate that if the number of required skills for computer science graduates continues to increase, degree programs might have severe difficulties in following this change. In addition, Krstić [19] performed research on big data

analysis to bring valuable business insights in the financial industry, indicating that methods can be used for any other domain to extract information. Moreover, the outcomes of these research studies are used for improving curricula, training, and performance management.

RESEARCH METHODOLOGY

DATA COLLECTION

The first step in collecting data is to select the relevant source of job advertisements. Here, *Indeed*, an American worldwide employment-related search engine for job listings was used [20]. Such websites offer a low cost of posting job advertisements which enables the employers to post more detailed descriptions of their requirements [21]. For the purposes of this study, 16 cities with the most numerous job advertisements were collected. Most of them are from the USA, several from the UK, and one city from Ireland and China. Thus, around a thousand job advertisements were collected for every city, over the period from January 19th 2020 to March 23rd 2020.

CONTENT ANALYSIS

Content analysis is a research method for studying documents and communication artifacts, which might be texts of various formats, figures, audio, or video used to examine patterns in communication in a replicable and systematic manner [22-24]. There are numerous examples of research based on content analysis that shows how text mining is used in various areas [25, 26]. A key element of content analysis is to code data in a way that will categorize it [27]. To accomplish this, the content analysis uses variables to represent the counts or proportions of keywords encountered within the records of text [28]. Using content analysis, four key parts of job advertisement are analysed: *Job Title*, *Company*, *Job Description*, and *Salary*. Figure 1 illustrates the process of data collection, content analysis, and presentation of the results.

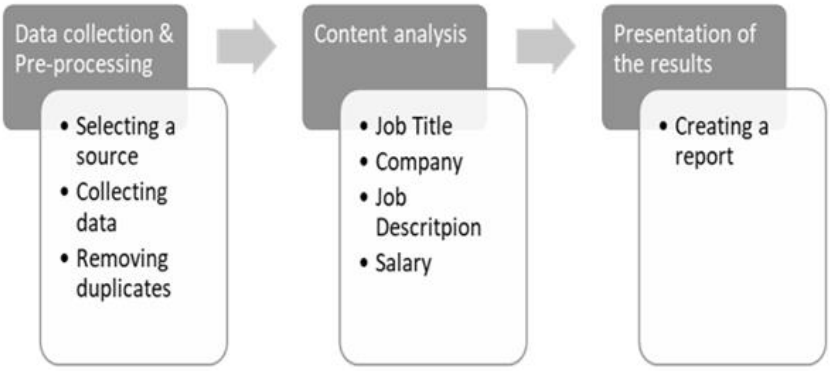

Figure 1. The process of data collection, content analysis and results presentation.

RESULTS

JOB TITLE ANALYSIS

According to UK Office for National Statistics, during the period from January to March 2020, the number of vacancies in professional, scientific, and technical activities, administrative and support service activities, accommodation and food services, human health, and social work, as well as in wholesale and retail trade increased significantly [29]. From Table 1, it is

evident for Birmingham, Manchester, Edinburgh, and London that these facts are confirmed for the majority of sectors. The vacancies are most numerous in administrative and support service activities. On the other hand, vacancies in professional, scientific, and technical activities are represented but to a much lesser extent than projected.

Table 1. Frequencies of job vacancies in the cities worldwide that appeared on *Indeed* from January 19th, 2020, to March 23rd, 2020 (continued on p.516).

Vacancy	Cities*									
	Bir	Man	Edi	Lon	Dub	NY	LV	Bos	SF	HK
Administrator	66	72	60	54						9
Administrative Assistant	17	30	11	30	8	27	10	3		20
Analyst Operations					5					
Advisor						10		2		
Business Analyst					3				3	10
Buyer					6				3	
Cashier						10	28	7		
Cleaner	15	11	14		5	15				
Collections Specialist					3					
Compliance Analyst					4					
Content Review Associate					8					
Consultant						13	12			12
Customer Adviser	20		14	12	4					
Customer Assistant	23	18	14	33						
Customer Care					3					
Customer Service	13		38							
Customer Specialist					3			2		
Customer Support					3					
Data Analyst					6				7	3
Delivery Driver			14		4					
Designer			13		3			2		5
Development Representative					3					3
Director of Digital Sales					3					
Dishwasher					20			3		
Executive Assistant					5					9
Finance Associate									4	
Floor Assistant					3	20		2		
Graphic Designer								5	8	3
Housekeeper							14			
HR Assistant						12				
Investigator						22		2	9	
Makeup Artist					3					
Operations Assistant									12	11
Personal Assistant				14						
Porter							12	2	4	
Receptionist						23	14			12
Retail Assistant					3		12		7	
Sales Advisor					3		48			5
Sales Assistant		18		18		15				
Sales Associate	17	12		10	6		14		5	

Table 1. Frequencies of job vacancies in the cities worldwide that appeared on *Indeed* from January 19th, 2020, to March 23rd, 2020 (continuation from p.515).

Secretary Assistant					3			7		
Security Officer							16	2	8	
Services Representative							10		7	
Settlements Analyst					3					
Software Developer					4					
Software Engineer					5				3	
Support Officer		10		10						
Support Worker		10	33							
Teacher							12		3	8
Technical Support					6				3	
Transporter						10				
Warehouse Operative	14		11				12		3	

*Bir – Birmingham, Man – Manchester, Edi – Edinburgh, Lon – London, Dub – Dublin, NY – New York, LV – Las Vegas, Bos – Boston, SF – San Francisco, HK – Hong Kong

Recently, it was estimated that demand for occupations in professional, real estate, scientific and technical activities would see an increase by 425 000 jobs by 2041 [8]. Information and communication, and education, health, and social work, and support service activities are also expected to increase in vacancies. Administrative and secretarial occupations are the only occupation group expected to see a London-wide decline in demand [30]. However, these predictions seem not quite accurate, although there is still a large period left to see if demand for administrative occupations will decrease as it is expected. These results only indicate that vacancies for the observed occupations were formally published. The demand for occupations from other mentioned sectors might also be high despite obtained results as employers may use other ways to fill the vacancies.

For Dublin, Ireland, the only sector with a marked increase in job vacancies since 2008 has been in professional, scientific, and technical activities [31]. Occupations like software developer, data analyst, and software engineer confirm that demand in this sector exists. Job vacancy rate in administrative and support activities, accommodation and food activities, information and communication and education activities stagnates during the last two years [31].

The US Bureau of Labor Statistics made a projection of trends in the labour market for the period 2018 to 2028. A slight increase in job vacancies is expected in the service-providing sector. However, sectors that are expected to see the fastest employment growth are health care and social assistance, private educational services, and construction. An increase in job vacancies is also expected in computer and mathematics and in renewable energy fields. On the other hand, several sectors are expected to see a decline in employment: retail trade, wholesale trade, utilities, the federal government, and manufacturing. A reason behind a decline in employment in retail trade is a shift to e-commerce [32]. This might also contribute to decreasing in job vacancies for the transportation and warehousing sector as well.

For New York, the most represented occupations are from the service-providing sector, as was expected. However, there are several deviations from the facts presented by the US Bureau of Labor Statistics. Occupations in transportation, retail trade, and wholesale sector are still among the most frequent vacancies which might indicate that the more significant decline in employment in these sectors has not occurred yet [33]. It is also noticeable that occupations such as administrative assistants are constantly in demand.

In Las Vegas, tourism is the biggest industry in Las Vegas and the major source of income. Therefore, the highest demand is expected to be in the service-providing sector. The results of

the analysis confirm these expectations, showing that most of the most frequent vacancies are exactly in the service-providing sector. Furthermore, the occupations in administrative activities, are in much less demand considering previously analysed cities.

For Boston, occupations that occur more than ten times are still mostly from the service-providing sector. However, occupations such as illustrator, copywriter, graphic designer, videographer, and animator are also represented which leads to the conclusion that the occupations of the creative industry are high in demand in Boston. As was the case in most previous examples, occupations in administration are rich in vacancies.

For San Francisco, the distribution of vacancies by sectors is much more diverse than in other cities. Job vacancies indicate that occupations in information technology, marketing, sales, science, engineering, and business management are in demand.

Overall, the results from Table 1 confirm that similar trends were identified in the USA as well as in the UK and Ireland. The most announced occupations are in administration and service-providing activities. The example of Las Vegas shows that demand is higher for the occupations in the industry which city is known for. Although a large increase in employment in healthcare, science, technology, and engineering is projected, occupations in these industries are not so represented in given results which might implicate that there are other ways of finding employees for the positions in these industries.

For Hong Kong, the most common vacancies are in consultancy, finance, and administration. Also, Hong Kong has a high demand for administrative assistants, which is a trend that was already seen in the USA, Ireland, and the UK. Occupations such as designer, graphic designer, and copywriter suggest that growing demand for professionals in the multimedia and creative industry exists. Also, vacancies for teachers appeared as projected. Language teachers and skill teachers like coaches are the most sought-after occupations of this type in Hong Kong [34].

ANALYSIS OF THE EMPLOYERS

The most common employer in the UK is NHS Scotland, which occurred in 72 job advertisements. According to the Office for National Statistics, NHS Scotland is a public sector body, and it is among the top 10 largest public sector employers in the United Kingdom. Other public bodies listed above are the University of Manchester, Manchester University NHS Foundation Trust, and the University of Edinburgh. In addition, public listed companies in the retail sector such as Tesco, boohoo.com, and Morrisons are also among the most common employers. The private company Iceland also belongs to the retail sector. It should be also noted that several recruitment companies are among the most common employers, such as Pertemps Network Group, Page Personnel, and Web2Recruit. All these companies are private. Finally, Sykes is another public company specialized in customer service and technical support. To summarize, different types of employers use the website Indeed for finding employees who will fill job vacancies. Mostly public sector bodies and public listed companies. Private listed companies are slightly less common among employers. There are also examples of charitable sector organizations such as The Action Group from Edinburgh. The public sector bodies are mainly from the health and education sector. Also, most of the employers are from the retail sector which confirms the growing demand for occupations in retail trade and wholesale.

The results of the Dublin job vacancy analysis suggested a growing demand for occupations in professional, scientific, and technological industries, so IT and technology companies are expected to be among the most common employers. IT companies such as Facebook, Google, Accenture, Salesforce, and Microsoft are in the top 10 most represented employers. It should be noted that the most represented IT companies are American. However, Accenture is an

example of an Irish company and according to Forbes magazine, it is in the top four largest companies in Ireland. As was the case with the UK, recruitment agencies in Ireland also use Indeed for finding employees. The Hays company is an example of a recruitment agency. On the other hand, public sector bodies are not among the most represented employers on Indeed for Ireland. The only example of a government agency is a Health Service Executive from the healthcare sector. Another issue that suggests that labour market in Ireland is slightly different than the one in the UK is that employers from the retail trade sector are also a lot less represented among the most common employers. The only example here is the supermarket chain SuperValu. Lastly, there are examples of pharmaceutical companies such as Novartis, companies specialized in investment banking such as CITI, telecommunications companies such as 3 Ireland, and management consulting company Deloitte. To conclude, the analysis of the employers in Dublin confirmed several previously derived conclusions such as the growing demand for occupations in the technology industry as well as a more diverse structure of the labour market than in the UK. It is important to mention that this analysis covers only Dublin therefore, the situation in the whole of Ireland might be different.

The United States Federal Government is the biggest employer in the USA since six out of ten most represented employers are from the public sector and are funded by the USA government. Similar trends found in the UK are repeated in the US, where universities are among the most common employers. However, unlike the UK where the employers from the retail sector are highly represented in the top ten most common employers, Nordstrom is the only example of the retail industry in the top ten most represented employers. Furthermore, NBCUniversal and Marriott International are examples of the entertainment and hospitality industry, which were not represented in the previously observed countries. Also, unlike Ireland, IT companies are much less represented, with only one example, Apple. It is important to add that the second largest employer in the USA, Walmart, occurs in only three job advertisements. One of the possible reasons might be that Walmart is using different platforms for announcing job vacancies.

To sum up, most of the employers in the USA belong to the public sector bodies, especially to the education sector. Other employers belong to various industries, from hospitality and technology to the media and retail industry. The results showed that employers from various industries use Indeed to announce job vacancies. On the other hand, some of the largest employers in the USA were not found among the most represented employers such as Walmart or Kruger which might suggest that there are other popular platforms besides which are used by these companies or there are other ways of finding employees.

Unlike other observed cities, most of the employers in Hong Kong are private companies. The single example of the public sector among the most common employers, the Hong Kong Productivity Council, may lead to the conclusion that Indeed is not a primary platform for announcing job vacancies in the public sector in Hong Kong. Furthermore, Earth.org is an example of a non-government organization that provides a scientific understanding of climate change and promotes environmental protection. As was the case in the UK, recruitment companies are among the most common employers on Indeed. Examples of the recruitment companies in the Table above are Classy Wheeler and Zebra Strategic Outsource Solutions. Companies specialized in logistics are Asia Airfreight Terminal and Cathy Pacific Services. Employers from the financial services and banking industry are highly represented. HSBC, CITI, Standard Chartered, and China CITIC Bank International are examples that illustrate the previous point best. Lastly, ITCS Group is an example of the IT industry which is the most represented in job advertisements. Thus, most of the employers presented on Indeed are private companies from the financial services and banking industry. Employers from the retail sector are not among the most common employers as was the case in the UK.

JOB DESCRIPTION ANALYSIS

Employers have been found to require a broader set of skills over time and these skills are no longer closely related to the occupation [10], rather to social and personal characteristics. It has also been found that employers require a wide range of skills, even for low-skilled occupations as well as for new and possibly high-skilled occupations [11].

While examining job advertisements, it was found that employers tend to focus on company branding instead of skills they require from candidates. This must be taken into consideration during analysis and presentation of the results.

Types of words have a major role in understanding the required skills of the employees. Research performed by Seljan [33] shows that verbs are usually used for describing duties, they create multi-word units consisting of verb phrases (VPs), which appear less frequently than noun phrases (NPs). NPs (noun preceded by adjective or noun) appear the most frequently carrying the meaning of the sentence, in the domain of law, or when extracting terminology using language-independent methods in the medical domain using the statistical and hybrid approach for information retrieval purposes [34].

While examining gathered job descriptions, it was found that the types of words employers use in the *Job Description* section can reflect whether employers focus on the duties of the candidates, or they prefer to list the required skills. As it was mentioned before, employers who use the *Job Description* section for branding companies must be taken into consideration. For the analysis, the first step was removing stop words to shorten the processing time. The next step was tokenizing the words and categorizing them by type.

The share of word types in job advertisements for the USA is given as follows; adjectives 8,8 %, verbs 8,9 %, nouns 38,5 %, other 43,8 %. The UK has a very similar division of the shares of word types to that of the USA; adjectives 8,3 %, verbs 9,3 %, nouns 35,1 %, other 47,3 %. Verbs are usually used for describing the duties of employees or for describing what does the company does. The share of adjectives that is almost like the share of verbs might suggest a growing trend of describing demanded skills. The share of word types in job advertisements for Dublin is given as follows; adjectives 8,8 %, verbs 8,2 %, nouns 35,5 %, other 47,5 %, while for Hong Kong is: adjectives 8,4 %, verbs 7,3 %, nouns 39,4 %, other 44,8 %. The share of adjectives is bigger than the share of verbs. The reason for this might be that employers are more focused on the skills they require from the candidates.

Skills analysis is based on the competence model 'KODE' [35], that is competencies are divided into four fields: Personal competence (P), Decision-making and responsibility (A), Professional and methodical competence (F), and Social and communicative skills (S). Each field consists of four blocks and each block contains four competencies. The results are shown in Table 2 as percentages of job advertisements in which competencies from the category appear.

The first aspect to point out is that three out of four observed countries and cities have the largest share of social and communicative skills represented in the job advertisements. This

Table 2. Skills by categories found in job advertisements

Skills	USA	UK	Hong Kong	Dublin
Personal Competence (P)	11 694	9 077	11 781	9 681
Social and communicative skills (S)	35 534	13 649	4 521	9 980
Decision-making and responsibility (A)	4 426	5 206	4 521	5 589
Professional and methodical competence (F)	12 476	9 009	4 110	8 383

confirms the outcomes of the study conducted by Dörfler and Werfhorst [10]. According to the *Indeed* career guide, skills such as verbal and written communication, teamwork skills and problem-solving skills, are ultimate skills that are demanded no matter occupation.

The results of the analysis shown above confirm that these skills are extremely important to employers. In addition, there is also a high demand for candidates who are customer-oriented and who possess conflict handling skills as well as language skills. All of this points to the conclusion that employers are looking for candidates who have solid relationship management skills and who work great with both clients and other employees.

The second most represented category is personal competence. For Hong Kong, it is the most represented category. As it is desired to have excellent interpersonal skills, according to employers, having excellent self-management skills is important as well. Motivation is a highly sought-after competency, which indicates that employers look for candidates who are focused on their professional growth. Responsibility, flexibility, and willingness to learn stand out as the highly represented competencies in this category. Employers also cite self-discipline, credibility, and reliability as desirable skills from this category. This highlights the fact that employers seek individuals they can rely on.

It is interesting that professional and methodical competence is the third most represented category. However, skills that stand out in this category as the most required are time-management skills and organisational skills. Employers highlight that employee must be able to complete tasks in a timely manner. Work ethic is also found as highly demanded by employers. Attention to detail and presentation skills are mentioned by many employers as highly desirable skills. From occupation-related skills in this category most required are analytical skills, products knowledge, and project management. It should be noted that this category is the least represented category in Hong Kong.

Finally, the last category is decision-making and responsibility. Although this category is the least represented, there are several sub-categories that are found highly demanded such as leadership skills and stress handling skills. Enthusiastic and results-driven candidates are found to be desirable to employers as well. This indicates that employers seek positive individuals as the energy of an individual can affect the entire team. In conclusion, the proportion of ads that have keywords associated with the skills categories reflects the relative importance of these skills for employers [21].

Numerous occupations require a certain level of education to be eligible to work in that profession. For example, New York, Miami, and Washington are cities with the highest percentages of occupations requiring a degree. As expected, the highest demand is for a bachelor's degree. However, most job advertisements do not require any degree. On the other hand, low percentages for the UK indicate that employers in the UK usually do not state the level of education they require or most occupations in job advertisements do not require a degree. It is the same case with Hong Kong as well.

Regarding salaries, the job advertisements which provide information about the salary are in the minority. Job advertisements from Dublin and Hong Kong did not provide enough information about the salaries. It was found only that the job of an administrative assistant is usually better paid in the USA with the average salary in the range from $40 000 to $60 000 while in the UK average salary for an administrative assistant is in the range from $20 000 to $30 000. According to one source which examines the salaries of the jobs in the USA, the median salary for the job of service representative in 2018 was $33 750.

DISCUSSION

The development of science and technology brought new ways of advertising job vacancies [9]. Advertising job vacancies is nowadays possible at a much lower cost [21]. This allows employers to write very detailed job advertisements with much more requirements and therefore it is getting harder for individuals to remain competent.

The content analysis proved to be a very efficient method in extracting meaningful information from job advertisements or any text in general [19]. The content analysis implemented and presented in this study derived useful insight into the employment trends for the USA, UK, Dublin, and Hong Kong. Useful conclusions were given about the skills that employers require, which are subject to constant changes.

From the results of this study, it is safe to conclude that there is a growing demand for social and personal skills than for professional and decision-making skills. Communication and organizational skills, strong work ethic, ability to adapt, motivation, and time-management skills proved to be essential skills employers require regardless of the industry or occupation. The analysis yielded occupations from the administration and service providing sector as the most demanded occupations in all analysed countries and cities. However, the share of occupations from other sectors in the job advertisements differs from city to city.

Although the results of the content analysis showed similar trends in the UK and the USA, the difference in skills demands between these countries exists. Only skills from the decision-making and responsibility category are equally demanded across these countries. A similar trend is noted in the case of Dublin and Hong Kong. However, it is also noted that skills from personal competence are equally demanded across these cities. There are several possible reasons for these results. Employers across different countries and cities might have different priorities when writing job descriptions. They can be focused on listing duties which candidates will perform or they can focus on the skills that candidates must possess. Also, there are employers who focus on promoting their company rather than requirements. The results also depend on the representation of the occupations since different occupations require different categories of skills. For example, occupations from the retail sector will probably require more social and communicative skills than occupations from the IT sector which will require more professional and methodical skills.

Another conclusion that emerged from the analysis is that employers find it crucial for candidates to have developed interpersonal skills. Equally important is to have excellent self-management skills meaning that the candidate must be motivated, willing to learn and grow in a professional way. All that points to the conclusion that employers are usually not focused on occupation-related skills. However, it is important to mention that this does not mean that occupation-related skills are not represented at all in the job advertisements. Occupation-related skills are demanded as well, although to a lesser extent.

The analysis of the employers showed that it depends on country to country which types of employers use online recruitment websites as a way of filling job vacancies. For instance, public sector bodies in the UK use recruitment websites unlike public sector bodies in Hong Kong. Finally, the analysis of the salaries gave insight into how salary for the same occupation differs from country to country or from city to city.

There are certain limitations of this study related to the quality, reliability, and representativeness of data from online job advertisements. These issues must be taken in consideration when using this kind of source as research data. For example, uncontrollable variables, like the quality of writing in job advertisements, can affect research [22]. In addition, the description of the job depends on recruiters' ability to communicate through written language [23]. Therefore, job

advertisements can possibly be ambiguous and hard to analyse. This also leads to challenges in coding this kind of problem. Furthermore, job advertisements usually reflect an ideal future state rather than the current reality of the labour market. The current demand on the labour market reflects development in a particular sector and does not necessarily represent the existing structure of the labour market. IT sector is the most suitable example to illustrate this point. IT sector has recently expanded and as a result, many countries record an increase in job vacancies in an IT industry. However, this does not necessarily reflect the actual share of the IT industry in a national structure of employers or employees [11].

Another challenge also occurs while using online job advertisements as a data source. Difficulties might occur while ascertaining whether the set of online job vacancies is a representative sample of all job vacancies in a specified economy. An example of this is international firms which usually first employ internally available candidates before announcing a vacancy in a labour market. In addition, in smaller towns or villages there are closer relations between populations, which often means that jobs are first offered to candidates known personally to the employer. This leads to the conclusion that jobs can be differently distributed and a part of them does not require a formal 'vacancy' announcement. This also indicates that online job advertisements can provide insight into what types of jobs employers find difficult to fulfil through internal or informal ways of finding candidates [11].

Thus, using job advertisements from an established portal and interpreting the results with caution can be a valid and acceptable choice. In addition, despite the previously mentioned limitations of job advertisements as research data, research using this source has been published in leading social science journals, suggesting that the field is expanding. Moreover, with the spreading of the Internet, reliance on Internet-based recruitment will possibly increase [9].

Another limitation of this study relates to the content analysis method itself. Although this is a very convenient method and it has numerous advantages for this kind of research, it has some disadvantages as well. The advantage of the human reading of job advertisements is that it may ensure that words are analysed in terms of their context as well as their frequency. Another disadvantage is that manual coding of an extensive data set is time-consuming. This may lead to the problem of biased or inconsistent coding, particularly if data are not cross-coded effectively.

Thus, future research might include more sophisticated methods applied in the domain of natural language processing, such as deep learning or other machine learning algorithms to extract meaningful patterns from the data. In addition, future avenues of this research might consist of a more detailed interpretation of the results represented in this article, e.g., further analysis could be also performed on the same data sample used in this article to derive additional conclusions.

Overall, our findings have several implications for job advertising strategists and can help marketers and managers understand how to structure communication content in such a way that it avoids common problems such as mismatch in terminology between requirements of industry and offer of educational and training institutions, etc. Managers can be guided by this research in deciding which characteristics of content to promote to elicit favourable responses among potential employees. Moreover, the outcomes of this research study can be used for improving curricula, training, and performance management.

REFERENCES

[1] Harper, R.: *The collection and analysis of job advertisements: A review of research methodology*.
Library and Information Research **36**(112), 29-54, 2012,
http://dx.doi.org/10.29173/lirg499,

[2] Maier, J.L.; Greer, T. and Clark, W.J.: *The management information systems (MIS) job market late 1970s-late 1990s.*
Journal of Computer Information Systems **42**(4), 44-49, 2002,
[3] Gardiner, A; Aasheim, C.; Rutner, P. and Williams, S.: *Skill Requirements in Big Data: A Content Analysis of Job Advertisements.*
Journal of Computer Information Systems **58**(3), 1-11, 2017,
http://dx.doi.org/10.1080/08874417.2017.1289354,
[4] Backhaus, K.: *An Exploration of Corporate Recruitment Descriptions on Monster.com.*
Journal of Business Communication **41**(2), 115-136, 2004,
http://dx.doi.org/ 10.1177/0021943603259585,
[5] Baravalle, A. and Capiluppi, A.: *It jobs in UK: Current trends.*
Proceedings of the 3rd IEEE International Conference on Computer Science and Information Technology (IEEE ICCSIT), Chengdu, July 2010,
[6] Smith, D. and Ali, A.: *Analyzing computer programming job trend using web data mining.*
Issues in Informing Science and Information Technology **11,** 203-214, 2014,
http://dx.doi.org/10.28945/1989,
[7] Surakka, S.: *Analysis of Technical Skills in Job Advertisements Targeted at Software Developers.*
Informatics in Education **4**(1), 101-122, 2005,
http://dx.doi.org/10.15388/infedu.2005.07,
[8] Molinero, C. and Xie, A.: *What do UK employers want from OR/MS?*
Journal of The Operational Research Society **58**(12), 1543-1553, 2007,
http://dx.doi.org/10.1057/palgrave.jors.2602286,
[9] Lee, S. and Lee, C.: *IT managers' requisite skills*.
Communications of the ACM **49**(4), 111-114, 2006,
http://dx.doi.org/10.1145/1121949.1121974,
[10] Dörfler, L. and Werfhorst, H.: *Employers' demand for qualifications and skills.*
European Societies **11**(5), 697-721, 2009,
http://dx.doi.org/10.1080/14616690802474374,
[11] Kureková, L.M.; Beblavý, M. and Thum-Thysen, A.: *Using online vacancies and web surveys to analyse the labour market: a methodological inquiry.*
IZA Journal of Labor Economics **4**(1), 2015,
http://dx.doi.org/10.1186/s40172-015-0034-4,
[12] Pejić Bach, M.; Bertoncel, T.; Meško, M. and Krstić, Ž.: *Text mining of industry 4.0 job advertisements.*
International journal of information management **50**, 416-431, 2020,
http://dx.doi.org/10.1016/j.ijinfomgt.2019.07.014,
[13] Bennett, R.: *Employers' Demands for Personal Transferable Skills in Graduates: a content analysis of 1000 job advertisements and an associated empirical study.*
Journal of Vocational Education & Training **54**(4), 457-476, 2002,
http://dx.doi.org/10.1080/13636820200200209,
[14] Wade, M.R. and Parent, M.: *Relationships between job skills and performance: a study of webmasters.*
Journal of Management Information Systems **18**(3), 71-96, 2001,
http://dx.doi.org/10.1080/07421222.2002.11045694,
[15] Huang, H.; Kvasny, L.; Joshi, K.D.; Trauth, E.M. and Mahar, J.: *Synthesizing IT job skills identified in academic studies, practitioner publications and job ads*.
Proceedings of the special interest group on management information system's 47th annual conference on Computer personnel research, ACM, Limerick, 2009,
[16] Iyer, H.: *A profession in transition: towards development and implementation of standards for visual resources management. Part A - the organization's perspective.*
Information Research **14**(3), 2009,

[17] Payne, H.: *Information management: a contemporary study of the discipline and the profession.* Master's Thesis.
Economic and Social Studies, Aberystwyth University, Penglais, 2009,
[18] Reeves, R. and Hahn, T.: *Job Advertisements for Recent Graduates: Advising, Curriculum, and Job-seeking Implications*.
Journal of Education for Library and Information Science **51**(2), 103-119, 2010,
[19] Krstić, Ž.; Seljan, S., and Zoroja, J.: *Visualization of Big Data Text Analytics in Financial Industry: A Case Study of Topic Extraction for Italian Banks*.
ENTRENOVA - ENTerprise REsearch InNOVAtion **5**(1), 35-43, 2019,
[20] Wikipedia contributors: *Indeed.*
Wikipedia, The Free Encyclopedia,
http://en.wikipedia.org/w/index.php?title=Indeed&oldid=958828080, accessed 8th May 2020,
[21] Sodhi, M. and Son, B.G.: *Content analysis of OR job advertisements to infer required skills.*
Journal of the Operational Research Society **61**(9), 1315-1327, 2010,
http://dx.doi.org/10.2139/ssrn.1640814,
[22] Xu, H.: *The Impact of Automation on Job Requirements and Qualifications for Catalogers and Reference Librarians in Academic Libraries.*
Library Resources & Technical Services **40**(1), 9-31, 1996,
http://dx.doi.org/10.5860/lrts.40n1.9,
[23] Ahmed, S.: *Desired competencies and job duties of non-profit CEOs in relation to the current challenges: Through the lens of CEOs' job advertisements.*
Journal of Management Development **24**(10), 913-928, 2005,
http://dx.doi.org/10.1108/02621710510627055,
[24] –: *Content analysis*.
http://en.wikipedia.org/wiki/Content_analysis, accessed 8th May 2020,
[25] Pejić Bach, M.; Krstić, Ž.; Seljan, S. and Turulja, L.: *Text mining for big data analysis in financial sector: A literature review.*
Sustainability **11**(5), No. 1277, 2019,
http://dx.doi.org/10.3390/su11051277,
[26] Ćurlin, T.; Jaković, B. and Miloloža, I.: *Twitter usage in tourism: a literature review.*
Business Systems Research **10**(1), 102-119, 2019,
http://dx.doi.org/10.2478/bsrj-2019-0008,
[27] –: *Natural Language Toolkit.*
http://en.wikipedia.org/w/index.php?title=Natural_Language_Toolkit&oldid=951031114, accessed 27th May 2020,
[28] –: *NumPy*.
http://en.wikipedia.org/w/index.php?title=NumPy&oldid=958607142, accessed 28th May 2020,
[29] Office for National Statistics: *Vacancies by industry.*
http://www.ons.gov.uk/employmentandlabourmarket/peoplenotinwork/unemployment/datasets/vacanciesbyindustryvacs02, accessed 28th May 2020,
[30] Greater London Authority: *London labour market projections 2016*.
GLA Economic, London, 2016,
[31] Nugent, C.: *Labour Market Trends in the Republic of Ireland (2019).*
Nevin Economic Research Institute (NERI), 2019,
http://dx.doi.org/10.4324/9781315792521-2,
[32] Office of Occupational Statistics and Employment Projections: *Employment Projections: 2018-2028 Summary*.
US Bureau of Labor Statistics, Washington, 2019,
http://www.bls.gov/news.release/ecopro.nr0.htm, accessed 28th May 2020,
[33] Seljan, S. et al.: *Comparative Analysis of Automatic Term and Collocation Extraction.*
INFuture 2009: Digital resources and knowledge sharing, pp.219-228, 2009,
http://infoz.ffzg.hr/INFuture/2009/papers/INFuture2009.pdf, accessed 28th May 2020,

[34] Seljan, S.; Baretić, M. and Kučiš, V.: *Information Retrieval and Terminology Extraction in Online Resources for Patients with Diabetes*.
Collegium Antropologicum **38**(2), 705-710, 2014,

[35] Heyse, V. and Erpenbeck, J.: *Kompetenzmanagement – Methoden,Vorgehen, KODE(R) und KODE(R)X im Praxistet*.
Waxmann Verlag GmbH, Münster, 2007.

Interdisciplinary Description of Complex Systems 19(4), 526-541, 2021

THE EFFECT OF TOURISM OVERNIGHT STAYS ON CROATIA'S EXTRA VIRGIN OLIVE OIL PRICES AND MARKET POWER: AN EMPIRICAL STUDY

Zdravko Šergo[1,] *, Jasmina Gržinić[2] and Anita Silvana Ilak Peršurić[1]

[1]Institut of Agriculture and Tourism
Poreč, Croatia
[2]Juraj Dobrila University in Pula – Faculty of Economics and Tourism 'Dr.Mijo Mirković'
Pula, Croatia

DOI: 10.7906/indecs.19.4.6
Regular article

Received: 11 February 2021.
Accepted: 16 November 2021.

ABSTRACT

The objective of this study was to analyse the impact of positive externalities of international tourism demand on increasing the market power (MP) of an extra virgin olive oil (EVOO) wholesaler in Croatia. In the context of this article, the MP measures how close the wholesaler can set the actual price of EVOO to the maximum the retailer wants to pay. Our hypothesis explained how the additional demand of tourist consumers for EVOO could stimulate and increase the MP of the wholesalers. Here, it was important to remember that the EVOO market signals relatively asymmetric quality information about products that varies in certain ranges. The selected time-series span the weekly period from 2017 to 2019. We used the Toda-Yamamoto approaches of causality in the relationship between the EVOO price gap and tourism overnights, as well as the autoregressive distributed lag model (ARDL) bounds test for cointegration. For larger EVOO bottles (0,75 l and 1 l), there is unidirectional causality flowing from tourism consumption, which we presume originates from the tourism demand variable, to MP. There is a relevant bidirectional causality in the case of the 0,25 l bottle. Tourism in a purchased bottle of 0,5 l does not manifest any side-effect impact on MP. This pioneering study has investigated the relationship between the MP of EVOO wholesalers in Croatia and tourist demand. An inventive view has been adopted with regard to the theoretical concept of measuring MP, but also due to the steps towards the use of ARDL bound testing.

KEY WORDS

extra virgin olive oil, market power, tourism, ARDL model, Toda-Yamamoto causality

CLASSIFICATION

JEL: D49, Q13, Z33

*Corresponding author, η: zdravko@iptpo.hr; +385 5240 8328;
52440 Poreč, Croatia

INTRODUCTION

This article investigates the MP of EVOO, which in our opinion is associated with rising international tourist demand, at least until the Covid-19 pandemic, using the example of the Republic of Croatia's wholesale trade. We argue that purchasing power is enriched by tourist buyers, which in turn causes EVOO price rises, empowering wholesalers' position in the supply chain. International tourist demand has recorded high growth rates in recent years, which has increased the consumer potential for buying bottled EVOO in Croatia. Using the positive externalities, our hypothesis asserts that tourist consumers cause an increased MP for wholesalers.

The research on market power (MP) in the olive oil trade is relatively scarce and it is mostly centred on exports [1, 2]. Before exploring the connections between MP and the growth in tourism, in order to justify researching the topic in that direction**,** we should state that the MP of wholesalers has grown generally because of their sizes, economics of scale, and organisational advantage. Wholesale provides a genuine economic function in the European economy; from a value chain perspective, it is an input to almost all production processes (the highest share of wholesale inputs had found in food and beverages and the lowest in construction, see in [3]). The rise of wholesalers has been driven by an intuitive complementarity between their sourcing of goods from abroad and an expansion of their domestic distribution network to reach more buyers. Both elements require scale economies and lead to increased wholesaler market shares and markups, as explained in [4].

During the past two decades, Croatia has experienced rapid growth and the spread of supermarket chains; moreover, the number of shelves in tourist centres has multiplied. Several factors account for this spread of shelves along with the occurrences of new trade counters, including the penetration of international supermarkets (Lidl, Kaufland, Spar, etc.) in Croatia seeking for a new profit source, greater economies of scale and scope, and more efficient procurement and distribution systems. Furthermore, the country's well-development tourism industry, EU membership, and expected increasing per capita income have encouraged a new demand potential for EVOO products, which influenced in engrossing supermarkets.

Due to increased costs and safety risks associated with losses, family olive planters (although this topic is out of the focus of this article) have been coerced to deliver produced EVOO to the wholesaler rather than leaving store surpluses unsold.

The wholesaler, in the oligopoly market structure, poses the strong potential for double marginalisation (both stages mark up prices above marginal cost) and its associated inefficiencies (double deadweight losses), as well as the potential effect on price transmission throughout the chain at the horizontal structure level, i.e. from the food processing stage to the retailing stage [5]. The modern distribution channels (hypermarkets, supermarkets and discount stores) and their prices are major final absorbers of these EVOO prices, which buyers accept (or become the default due to boycotting).

Tourist buyers may distort extra-market sources of information about EVOO reputations, and it is very rare that there is repeated trade that serves a correctional purpose. This is especially the case when EVOO with a falsified declaration of content quality is involved. The analysis of Croatia's IPTPO panellists, the internationally recognised olive oil tasters in the county among the very few, indicate that one in three oils chosen randomly does not correspond to the category of EVOO; also according to [6] most of the imported oils from abroad do not meet the quality which is declared on the labels of EVOO bottle.

In relatively unknown products, such is exotic EVOO bottled in a small volume with a decent design, functions as a souvenir as opposed to a grocery item; that items may not ever signal required quality via the price level.

Food souvenirs are tangible reminders of a travel destination and play an important role in the hospitality and tourism industry [7]. Bottled EVOO as a souvenir is often a mandatory purchase guided by an instinct desire to possess a suitable symbol of a meagre but above all healthy Mediterranean diet. In this context, the organoleptic quality does not have to be commensurate with the price. Under such circumstances, the utility function can be non-negative for a specific price range and generate an inverted U-shaped function. According to [8], the EVOO market in Chile is a good example. Therefore, minimal and maximal prices can significantly deviate, hitting consumer surplus. Consumer knowledge of this product, we argue in this article is still limited, especially for foreign, myopic, and naïve tourist consumers.

In this article, we also argue that in circumstances of asymmetric information communication, 'money for value' or EVOO prices may vary in ranges. That range constructs an imaginary price gap for which the trajectory can deviate from the average price, which may in turn anticipate the wholesaler MP. In this article, the model will be put to the test for the very first time the hypothesis that demand generated by tourism has significantly stimulated the widening price gap in EVOOS trading, driving the MP of wholesalers higher and higher.

Therefore, we have given serious attention to the 'tourism and MP effect' nexus which affects demand for the bottled EVOO market. To that end, we have used a different approach to investigate the relationship among olive oil MP and tourism growth, namely an ARDL cointegration test developed by [9-10], as well as the specific technique of causality analysis given in [11], based on weekly data for the period between 2017 and 2019. This recent three-year period has been chosen because these have been the most prosperous years in terms of Croatia's tourism industry development.

Modelling the relationship function between the MP of EVOO from a wholesaler perspective and tourism demand is an important area of research. The price gap, i.e. differential prices for EVOO, will be instrumented in our objective to deliver the impact of international tourism overnight stays on MP. The rest of this article is organised as follows: in Section 2 provides a short literature review; Section 3 explains the theory and general model and constructs the estimation procedure (including the unit root test, the ARDL approach and Toda-Yamamoto causality analysis), as well as providing the data; Section 4 explains the empirical results; and, finally, Section 5 offers concluding remarks and outlines the policy implications.

LITERATURE REVIEW

The subject of MP in the olive oil trade has not been studied extensively, at least according to academic articles composed in the past. A comparative analysis of the MP of olive oil exporting countries indicates that exports are imperfectly competitive in the EU market, and that Italy has higher MP compared to Spain and Greece [2]. Other authors reached a similar finding as far as Italy is concerned while Tunisia has the lowest MP in relative terms [1]. Both articles place the research problem in the environment of economic comparative studies of the Mediterranean countries, using the same method to measure the exporter's MP, as in [12].

In papers written by other authors, there is an assumption that olive oil tourism based on a typical gastronomic product of internationally recognised quality boosts sustainable destination development [13]. Oleotourism, or olive oil tourism, is a practice that contributes to highlighting an emblematic resource for Europe, with a special emphasis on the countries of the Mediterranean basin, receiving the mayor number of consumers interested in this type of resource [14]. Links between the tourism demand and olive oil supply, in various perspective are brought to light in manny papers [15-19]. Some authors, like [20] have emphasised that the

price of EVOO is highly subject to variability, meaning the consumer is not able to perceive a price of reference. In a study addressing consumer misuse of country-of-origin labels conducted in Italy, the authors test whether there is a price differential associated with the country-of-origin information for EVOO [21]. As a high value commodity on the market, EVOO is a suitable target for fraudsters [22]. To address whether there is a relationship between the value of attributes based on the market price and on consumer utilities a hedonic price (HP) approach is combined with the actual consumer utilities from a real choice experiment (RCE) for EVOO attributes in a study produced by [23]. Various tastes and preferences, demand and supply, the rental position of stores, and other complexities dictate the range of price heterogeneities for the homogenous product, accentuating standard consumer theory. Furthermore, search frictions resulting from agents' imperfect information about sellers' prices explain the price dispersion in otherwise homogeneous product markets [24]. Other authors underline that the olive-oil based tourism loses momentum in a rural destinations, since many cultural cities are offering tourism experiences focused on highlighting the cultural and sensory content of gastronomy and typical productions such as EVOO [25-27].

THEORY, METHODOLOGY AND DATA

THEORY

The literature suggests a considerable range of MP measures based on industry concentration measures, entry barriers, and various indexes (for more one can see [28]). We will not consider the theoretical underpinning in that direction because we are dealing with a homogenous product based emerging MP rather than a firm or a sector. Otherwise, profitability data or EVOO returns is unfortunately lacking.

Our theoretical consideration is based on [29], which tells that MP and a large markup arise because there is a product (in our case EVOO) that customers happen to enjoy and are willing to pay a premium to obtain. The retailor, we assume, buys EVOO at storage from the wholesaler agent, with whom they communicate on a regular basis. There is a maximum price that the retailer will pay for bottled EVOO from the wholesaler, and that price depends on the quality of the oil itself, the convenience of the store location, and how pleasant his or her staff are to the final customers. If the retailor strategically does business with zero profit margin, minimum prices will be a frequent attractor to that store. The actual price that the retailor pays lies somewhere between those two. In the opposite case, one of the two subjects (retailor or wholesaler) would refuse to participate in the transaction of traded EVOO. Further, if we transpose a previous theoretical story on the wholesaler perspective, we have to turn the situation upside-down.

The consequence of this mental exercise provides us with the metrics of MP. However, there is a different way to conceive of and measure the MP of EVOO on wholesaler grounds. Rather than focusing on total economic profits from EVOO trafficking (as stated previously, that micro-segmented panel data is unfortunately unavailable), we could instead focus on the extreme price poles (maximum and minimum) considering that the minimum price is approximately equivalent to the value charged by wholesalers relative to their costs.

The MP of wholesalers is therefore a measure of how close the wholesaler agent can set the actual price up to the maximum according to which that agent will sell considering the operating cost. The marginal cost per item in this case is the actual price paid to the lowest price a wholesaler will accept. Therefore, even if we do not know the profits that wholesalers receive in such exchanges, we can measure the MP. The larger the trade, the more abnormal

the return yielded by the MP. This might be supported by asymmetric information communication between the tourist customer and the retailor.

We assumed that the swollen demand for EVOO due to increased overnight tourist stays in recent years has positively affected the growth of the MP for wholesalers.

The set general theoretical model depicts the narrative connection, and this is followed by the empirical verification.

$$MP = f(+ NIGHT), \quad (1)$$

where MP is as we suggested, streamlining, ground on the differential between maximal price and minimal price; that gap is a proxy for the MP and NIGHT represents international overnight stays, which on the other hand is a proxy for international tourism demand.

Such a formulation is in line with the aforementioned general consideration regarding the link between overnight tourist stays and the MP in the oil wholesaler's market consideration.

Accordingly, the main hypothesis of this study is as follows: an increase in overnight stays involving international tourists boosts the MP of EVOO from a wholesaler perspective.

ECONOMETRIC METHODOLOGY

ARDL cointegration and bounds tests

To analyse the long-term relationship between a set of variables, authors [10] suggested the use of an autoregressive distributed lag procedure or bounds test that does not require stationary pre-testing, and which can be used regardless of whether the variables are I(0), I(1), or mutually cointegrated, given that none of the series is I(2). Despite these relaxing circumstances, we have produced a verification to determine whether second-order integration in some time series exists by conducting a ADF and DF-GLS unit root test in order to eliminate further exercises with data that encompass some of the variables. It is highly probable that the results will direct us to employ an ARDL bound test [10]. Therefore, if those tests show that the individual time series variable is either I(0) or I(1), an analysis with that price of EVOO will continue with the bounds test. The ARDL model has a certain number of advantages over traditional methods of testing cointegration. Firstly, as specified before, this method puts less strain on unit-root testing. Secondly, we can simultaneously estimate the short-run as well as long-run relationship among the variables using the ARDL bound testing procedure. Our sample has 156 weekly observations; the characteristics of long-term regressed series do not lose significance due to a considerable amount of high-frequency data. In addition, the ARDL model takes care of endogeneity issues by adding lags for both dependent and independent variables in the model. Finally, the ARDL model can be converted into a twin unrestricted error correction model (UECM), including both short-run and long-run dynamics. The bounds test is based on the following UECM:

$$\Delta Y_t = const + \Sigma\beta\Delta Y_{t-1}k_i = 1 + \Sigma\gamma\Delta X_{t-1}k_i = 1 + \omega Y_{t-1} + \theta X_{t-1} + \varepsilon_t, \quad (2)$$

where Y_t denotes price gap measured in EVOO and X_t denotes a tourist overnight stay as a single input (as explained previously), both expressed in natural logarithms. An appropriate lag selection is based on the Schwarz Bayesian Criterion (SBC). The automated model selection process involves choosing the maximum lag for each regressor, which is set as 8 (because the data is weekly). The ARDL procedure allows for the possibility that the variables may have different optimal lags (after the searching process has ended), whereas this is impossible with

conventional cointegration procedures. The null hypothesis for no long-term relationship between the price gap and the tourism input variable is not rejected, by testing the *F*-statistic, when:

H_0: $\omega = \theta = 0$,

against the alternative

H_0: $\omega \neq \theta \neq 0$.

Instead of the conventional critical values, authors [10] proposed a bounds test for two sets of critical variables. The first set assumes that all variables are (0), and the other set assumes that all variables are (1). If the tested *F*-statistic (or Wald statistic) value lies below the lower bound critical value, then the null hypothesis of a non-existent cointegration relationship cannot be rejected; further, if it exceeds the respective upper bound critical value, the null hypothesis is rejected. If the tested *F*-statistic value falls within the lower and upper critical value bounds, inference is inconclusive. Furthermore, because of the potential existence of a trend in the series (if the former case is unable to identify cointegration between two series), estimations are completed to satisfy the unrestricted intercept and no trend case (as an auxiliary test). Model diagnostic checking is particularly significant in the sense that some important ARDL assumptions such as errors are serially independent and normally distributed. Estimations are completed using an ordinary least squares procedure alongside the White's test for cross-sectional heteroscedasticity-consistent standard errors, as well as a covariance matrix, appropriate serial correlation diagnostics (the Breusch–Godfrey LM test), and the Jarque–Bera statistic for the normality test. The graphical recursive CUSUM and moving sum (MOSUM) of the recursive residuals according to [30, 31] are applied to detect whether any autoregressive structure is integrated into the model. These tests have been employed to assess the parameter stability of the model.

Causality analysis

The initiation of proceedings within Toda-Yamamoto causality analysis occurs if the cointegration link between MP and NIGHT persists. Formally speaking, if a price gap and an international tourist overnight input as a metric of consumption are regressed against one another in levels, the resulting residuals essentially represent error correction terms. These terms measure deviations in the long-run equilibrium between the two series. Hence, the ARDL (1) can be re-parameterised after replacing Y_{t-1} and X_{t-1} with the lagged residuals:

$$\Delta Y_t = const + \alpha \, \mathrm{ECT}_{t-1} . \Sigma_{i=1}^{n} \rho \Delta Y_{t-1} + \Sigma_{i=1}^{p} \sigma \Delta X_{t-1} + \mu_t, \quad (3)$$

e.g., the error correction model via the two-step procedure of Engle and Granger.

These lagged residuals represent an error correction term, denoted in this article by ECT_{t-1}, which provides an insight into the speed of adjustment to a long-run equilibrium within a particular time frame from a change to one of the series. Furthermore, if the coefficient of ECT_{t-1} is statistically significant (by t-value), long-run causality is confirmed. ECT_{t-1} should be between 0 and 1 with a negative sign, which implies convergence of the system back to the long-run equilibrium position.

Additionally, the direction of cause and effect between the variables – testing the hypothesis that tourist demand measured by the number of overnight stays causes an increase in the MP of EVOO wholesalers – will be clarified by using the Modified Wald test (MWALD) recycled according to the Toda-Yamamoto (1995) procedure. The MWALD test skips obstacles and problems associated with the classic Granger causality test resulting from non-stationarity or cointegration between series. It is to be expected that the latter problems would cause a theoretical inconsistency in collision with the empirical performance of the classical Granger

causality test [32-34]. Unlike the Granger causality test, the Toda-Yamamoto (1995) causality test is performed using a standard VAR model in levels rather than first differences; this process minimise the risk of misidentifying the order of integration [35].

To apply the Toda-Yamamoto [11] version of the Granger non-causality test, we summarise the MP–NIGHT model in the following VAR system:

$$MP_t = \alpha_0 + \sum_{i=1}^{k} \alpha_{1i}\, MP_{t-1} + \sum_{i=1}^{dmax} \alpha_{2j}\, MP_{t-j} + \sum_{i=1}^{k} \delta_{1i}\, NIGHT_{t-1} + \sum_{i=1}^{dmax} \delta_{2j}\, NIGHT_{t-j} + \gamma_{1t}, \quad (4)$$

$$NIGHT_t = \beta_0 + \sum_{i=1}^{k} \beta_{1i}\, NIGHT_{t-1} + \sum_{i=1}^{dmax} \beta_{2j}\, NIGHT_{t-j} + \sum_{i=1}^{k} \eta_{1i}\, MP_{t-1} + \sum_{i=1}^{dmax} \eta_{2j}\, MP_{t-j} + \gamma_{2t}. \quad (5)$$

From (3), the Granger causality from $NIGHT_t$ to MP_t implies $\delta_{1i} \neq 0$ for all i; similarly, in (4), MP_t Granger-causes $NIGHT_t$ if $\eta_{1i} \neq 0$ for all i. The bi-variable model is estimated using seemingly unrelated regression (SUR), as in [35].

ABOUT THE DATA

In order to examine the relationship among the variables, the study used weekly time-series data from 2017-2019 for four different volumes of bottled EVOO. In this study, the MP proxied by price gap is the dependent variable. Tourism demand is proxied by overnight stays (NIGHT), and this is the only variable singled out as the independent variable. MP in our reconsidered theory of price differential variability provides the basis for its dynamic measurement. The MP marked with different volume labels – detailed later in the text – will distinguish various prices per volume: 0,25 litres, 0,5 litres, 0,75 litres, and 1 litre. Auxiliary variables (maximal and minimal prices) for calculation of MP were sourced from an overview of wholesale EVOO prices by week given in Kn per litre, given in [36].

International tourism overnight stays (NIGHT) sourced from Eurostat [37] are used as a measure of tourist demand, and in this study have been used as an alternative (substitutive term) to tourism consumption given that this could not be deduced from the Eurostat website on a weekly frequency. This procedure has been undertaken to match the same data consistently with the MP time series variable.

A few extra words would be good to mention here, in this data section.

The tourism overnight time series data is unfortunately in the form of quarterlies. Because of the ARDL time series methodology, the postulated requirement of designed studies is a disaggregation of their lower-frequency value to higher (or weekly unit) values. The tourism overnight stays data and the inherited seasonality in that data has enabled us to obtain weekly tourism overnight stays entries with a prudent time span length (2017week1–2019Qweek52) through temporal desegregation techniques, as explained by Chow-Lin (for more, see in [38]). All the variables used in this article come in their natural log form.

The data series for the selected economies contains 156 weeks. Hence, we suppose that idiosyncratic outliers and structural breaks may be hidden in the Data Generating Process (DGP) of our weekly time series. As pointed out above, an adequate technique to handle these handicaps is the ARDL bounds test approach.

EMPIRICAL RESULTS

UNIT ROOT TEST

Before conducting tests for cointegration, it is vital to ensure that the variables under consideration have not been integrated at an order higher than 1. In the presence of I(2) or higher variables, the computed statistics provided by authors [10, 39] are not valid according to [40]. Thus, in order to establish the integration properties of the series, we used quick ADF

and DF-GSL unit root tests to confirm that none of the series was I(2). Accordingly, both tests were at level and at first difference (if the level test was non-stationary), and the results are shown in Table 1.

Table 1. Unit root test (ADF and DF-GLS) test.

	Augmented Dickey–Fuller Test		DF–GSL
Levels	MP1	–11,800 (0)***	–4,022 (1)**
First difference		–	–
Levels	MP075	–7,298 (1)***	–7,185 (1)***
First difference		–	–
Levels	MP05	–3,642 (3)***	–4,413 (2)***
First difference		–	–
Levels	MP025	–5,480 (2)***	–7,068 (1)***
First difference		–	–
Levels	NIGHT	–2,511 (4)	–2,912 (1)*
First difference		–4,988 (4)***	–5,021 (3)***

Note: all the regressions include a linear trend in the levels and include an intercept in the first differences.
Note: the numbers in parentheses are the optimal lag orders and are selected based on Schwarz Bayesian.
Note: the DF–GLS test statistics include an intercept and a linear time trend in the levels and only an intercept in first difference [41].
*significant at level 1 %,
**significant at the level 5 %
***significant at the level 10 %

As indicated in Table 1, the ADF and DF-GLS tests reject the null hypothesis of the unit root at levels for MP variables in the full spectrum, implying that the dependent variables are stationary at levels. Nonetheless, regarding the independent NIGHT variable, we do not accept the 10 % significance. The null hypothesis is however rejected at the 1 % significance level, at the first difference, implying that the NIGHT variables become stationary at the first difference. In our case, we have a mix of I(0) and I(1) variables. Therefore, we can apply the tests proposed by authors [10] and proceed with the test for cointegration.

RESULTS OF THE ARDL COINTEGRATION TESTS

In the first step in applying the bounds test, we specified the optimal lag length of the UECM, i.e., Equation (1), and checked the long-run level equilibrium relationship.

We have attempted to optimally set up the ARDL model and fixed an optimal lag, which is crucial. With an initial lag of 8, the automated model selection, according to minimal SBC [9], calculates the optimal lag length. The results of the cointegration test using ARDL are presented in Table 2. The estimated ARDL model has passed several diagnostic tests, meaning there is no evidence of serial correlation, heteroscedasticity, or deviation from normal distribution. The cumulative sum of the CUSUM and moving sum (MOSUM) for the recursive residuals plots – which is shown in Figure 1-4 from a recursive estimation of the model – also indicate stability in the coefficients over the sample period in all cases (in Appendix). This indicates that the model is well founded and suitable for the study of cointegration among the variables.

The F-statistics for cointegration analysis based on the selected ARDL models are reported in Table 2 for all EVOO price gap cases. All the reported F-statistics – as well as the t-statistics – lie above the upper bounds; consequently, the null hypothesis of no cointegration is rejected and the precondition for cointegration is established in all four EVOO volumes.

Table 2. Result of the cointegration test using ARDL Approach (case III).

Dep. variable	Indep. variables	Bounds F-test statistic	Bounds t-test	Cointegration	LM-test	JB-test	HET
MP025	Night	40,168 (0,000)	–10,967 (0,000)	Yes	0,285	0,548	0,855
MP05	Night	6,951 (0,001)	–4,485 (0,001)	Yes	0,031	0,643	0,620
MP075	Night	38,444 (0,001)	–10,725 (0,000)	Yes	0,989	0,496	0,879
MP1	Night	39,433 (0,000)	–10,849 (0,000)	Yes	0,941	0,924	0,937

Note: the critical values for the *F*-statistic are derived from table CI (III) (see Table 4). The range for the associated t-test is from lower-bound I(0) = –3,43 to upper-bound I(1) = –3,82.

Note: LM is the Lagrange multiplier test for serial correlation with a χ^2 distribution, with only one degree of freedom; JB is the Jarque-Bera test for normality; HET is the Whitetest for heteroscedasticity with a χ^2 distribution with only one degree of freedom.

Table 3. Critical Values for the ARDL modelling approach in relation to the Bounds test [10].

	Case III	
	I(0)	I(1)
10 % critical value	3,17	4,14
5 % critical value	3,79	4,85
2,5 % critical value	4,41	5,52
1 % critical value	5,15	6,36

SHORT-RUN ESTIMATES

Table 4 indicates the short-run implications of tourism overnight stays on the price gap, e.g. MP of wholesalers. While the EVOO price gap is found to have a lagged impact on itself in the case of 0,25 l (one lag) and 0,5 l (two lags), whereas in the case of the other two bottle profiles (0,75 and 1 l) there is an instantaneous autoregressive short-run impact. The price gap (or MP) affects its own trajectory negatively at a statistically significant level in two cases (in a bottle: volume 0,5 l and 1 l).

Overnight stays have a positive and statistically significant lagged impact on price gap, e.g. the MP of wholesalers, in all bottle cases except for the 0,75 l case where that impact is enhanced by the instantaneous die out. One time-lagged error correction term is negative and statistically significant at a 1 % level in the case of all the analysed price gaps for the various bottles.

LONG-RUN ELASTICITY

We consider weekly measurement of the included variables with 156 observations to be a sufficiently reliable basis to estimate the long-run interference of tourism overnight stays on the price gap in EVOO trading.

Table 4. Results of the short-run estimates.

Volume	Variables	Lags			
		0	**1**	**2**	**3**
0,25 l	ΔMP	–	–0,497 (–6,899)	–	–
	ΔNOC	0,001 (0,125)	0,097 (0,846)	0,023* (1,869)	0,507** (2,086)
	ECM_{t_1}	–0,811*** (–7,448)	–	–	–
0,5 l	ΔMP		–0,372 *** (–3,967)	–0,173** (–2,312)	
	ΔNOC	0,376 (0,126)	0,451 (0,024)	0,047** (2,231)	0,866* (1,935)
	ECM_{t_1}	–0,443*** (–4,599)	–	–	–
0,75 l	ΔMP	0,890 (1,315)	–	–	–
	ΔNOC	0,590** (2,153)	0,887** (2,332)	0,342 (1,276)	–0,414 (–0,743)
	ECM_{t_1}	–0,890*** (–10,823)	–	–	–
1 l	ΔMP	–0,920 (–10,850)	–	–	–
	ΔNOC	0,181 (0,845)	0,455 (0,873)	0,955 (0,821)	0,144*** (2,561)
	ECM_{t_1}	–0,870*** (–10,945)	–	–	–

*significant at level 1 %,
**significant at the level 5 %
***significant at the level 10 %

The long-run elasticity of a single independent variable (NIGHT) with respect to the dependent variable (MP) is shown in Table 4. All EVOO price gaps referring to various sold bottles have statistically significant and positive relationships, and are affected by an increase in international tourism overnight stays. Here, as predicted in the theoretical overview, we can state that tourism demand causes a price gap/MP increase for wholesalers.

Table 5. Long-run estimates of dependent variable MP.

Variables	0,25 l volume	0,5 l volume	0,75 l volume	1 l volume
NOC	0,76** (2,311)	0,756** (2,231)	0,329*** (2,897)	0,412* (2,119)
intercept	2,447 (1,452)	1,296 (1,397)	1,185 (1,264)	1,826 (1,003)

*significant at level 1 %,
**significant at the level 5 %
***significant at the level 10 %

TODA-YAMAMOTO CAUSALITY TEST

After estimating long-run results, we proceeded to the causality test. However, first the long-run causality is conducted by the coefficient t-statistics, which stand before the ETCt-1,

wherein this term measures how fast the deviations from the long-run equilibrium die out following changes in the NIGHT variable. The lagged ECT coefficients from Table 5(4) show that international overnight stays and MP in all types of contracted bottled EVOO were restored to the long-run equilibrium. The analysis of Equation (6), as presented in Table 7, indicates that there is long-run cointegration among the variables at the 1 % significance level; moreover, the ECT coefficients revealed in Table 8 imply that any deviation from the long-run equilibrium is corrected within the adjustment speed range of 44-89 % (the quickest speed is measured for the 0,75 l volume EVOO bottle). This also indicates a strong causality for the tourist demand variable on the MP of EVOO wholesalers.

From the estimation of the Toda-Yamamoto Granger causality test (see Table 5), we can make the following assertion based on the results of this study:

For 0,75 l and 1 l bottled EVOO there is unidirectional causality flowing from international overnight stays demand to the price gap (MP), thereby supporting our hypothesis that tourism affects MP in EVOO trading. For the 0,25 volume bottle, the reverse causality evidence, where tourism overnight stays is caused by price gap, revealed that the MP also has side-effects on tourism overnights, thus forming a bidirectional causality linkage.

Table 6. Toda-Yamamoto no-causality test two-variate VAR model results.

EVOO volume	Lag(k)	Lag (k+dmax)	Chi‑sq	Prob.	Direction of causality
0,25 l	1	1+2	5,654	0,098*	NIGHT → MP
	1	1+2	13,455	0,018**	MP → NIGHT
0,5 l	1	1+2	2,756	0,446	NIGHT do not cause MP
	1	1+2	9,966	0,813	MP do not cause NIGHT
0,75 l	1	1+2	12,734	0,013**	NIGHT → MP
	1	1+2	0,814	0,556	MP do not cause NIGHT
1 l	1	1+2	9,388	0,024**	NIGHT → MP
	1	1+2	0,896	0,403	MP do not cause NIGHT

Note: k+dmax denotes VAR order. The lag length selection was based on AIC: Akaike information criterion.

*significant at level 1 %,

**significant at the level 5 %

***significant at the level 10 %

→one−way causality

CONCLUSIONS AND POLICY IMPLICATIONS

This article was directed towards attaining a full understanding of the causal relationships between international tourism overnight stays and the MP of EVOO from the wholesaler's perspective. For the bottled EVOO products that pass the rigorous statistical testing – the unit root, cointegration, and bounds testing [10], Toda-Yamamoto causality analysis was also particularly important.

Having applied the ADF (Augmented Dickey-Fuller) and DF–GLS (Elliott, Rothenberg, and Stock) unit root tests, we found that the variables are a mix of I(0) and I(1). Hence, the ARDL bounds test approach of cointegration has been employed followed by the Toda- Yamamoto Granger causality test in order to shed more light on the relationship between international tourism stays and the MP of wholesaler agent proxies by price gap. The positive empirical evidence about cointegration is acquired from ARDL bounds test and research proceeded by investigating facts about the short-run and long-run relationship, as well as the ECM based strong causality test and finally the Toda-Yamamoto Granger causality test. The results from the ARDL

approach demonstrate that, regardless of which volume has been purchased, EVOO reveals cointegration, which is manifested in the assumed link between the MP and tourist demand.

There is significant statistical evidence that tourist demand has swelled in recent year, and that international tourism overnight stays have had a positive impact on the MP growth of EVOO in both the short-run and the long-run.

For larger bottles (0,75 and 1 l), there is unidirectional causality flowing from tourism consumption, which we supposed originated from the NIGHT variable, to the artificially constructed MP variable. This fact supports our thesis that tourism growth impact elevates the MP of wholesalers thanks to the product discussed within this article. The feedback hypothesis discovered here has been found to be relevant in the case of 0,25 l. In this case a bidirectional causality unfolded. In addition, 0,5 l bottled EVOO does not manifest any side-effect relationship at all, at least when this issue was assessed via Toda- Yamamoto causality testing.

Our insights are invaluable for the implementation of any relevant tax policy measures/instruments with the objectives of advancing either non-board consumption or augmenting the consumer surplus regarding EVOO purchases, presupposing the desire to pursue olive style tourism. The consumers, in particular foreign tourists, are not very knowledgeable; put bluntly, they are myopic regarding the attributes of EVOO and may misperceive the price as a proxy for quality. Our empirical results also suggest the following implications that public policy must touch on. The Croatian executive authorities should encourage a policy of reducing agrarian import dependence in general and support the production of EVOO in Croatia. The EVOO stored by wholesalers should be controlled and organoleptic properties should be analysed in order to eliminate the information white noise throughout the rest of the supply chain. Since the final consumers for these smaller or larger bottles of oil are predominantly tourists, the goal must be that foreign tourists have a better feeling about the price-quality ratio. That policy may reduce the MP trading with the EVOO within Croatia despite the tourist demand, which will perpetuate the same phenomenon on a nationwide scale.

There are certain limitations of this study. This study was focused on measuring wholesalers' market power. However, they are only one stakeholder in the observed supply chain. Therefore, we propose guidelines for further research in this insufficiently researched topic along with the line questioning how tourism demand impact MP of EVOO. The next research could examine other stakeholders in the olive oil supply chain in Croatia (marketing agency, transport, the state). Direct channels, the packaging cost, the processing subventions that boost supply, were not included in the research and they might be incorporated for future research. Although Croatia, as well as other Mediterranean countries, has a long olive oil production tradition, the production of olive oil in Croatia almost doubled in the last 20 years. Consequently, olive production surfaces have increased from 11,398 ha in 2000 to 18.606 ha in 2019 followed by an increase in olive oil production volume (from 28,188 hl in 2000 to 44,497 hl in 2019) [42]. In a future study, somehow detached from our central topic, it could be investigated, by the quasi-experimental method of difference-in-difference regression how the input of state subsidies affected the growth of olive oil production.

ACKNOWLEDGEMENT

This article is a result of scientific project Tourism development and destination impacts supported by the Faculty of Economics and Tourism 'Dr. Mijo Mirković', Juraj Dobrila University of Pula. Any opinions, findings, and conclusions or recommendations expressed in this article are those of the author(s) and do not necessarily reflect the views of the Faculty of Economics and Tourism 'Dr. Mijo Mirković' Pula.

APPENDIX

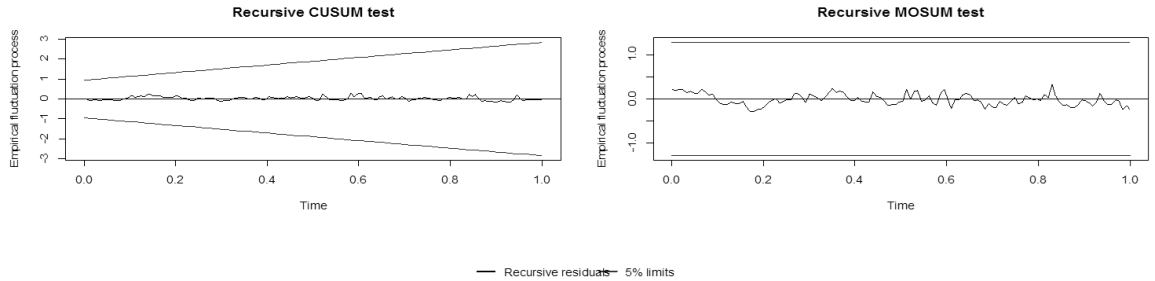

Figure 1. Plot of CUSUM and MOSUM: case of MP025.

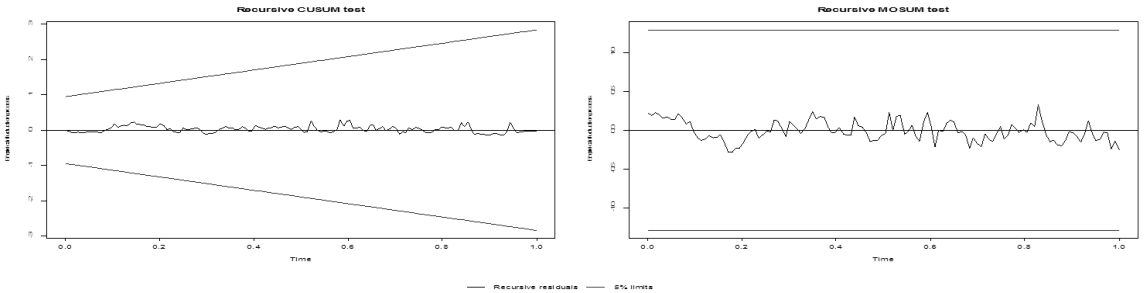

Figure 2. Plot of CUSUM and MOSUM: case of MP05.

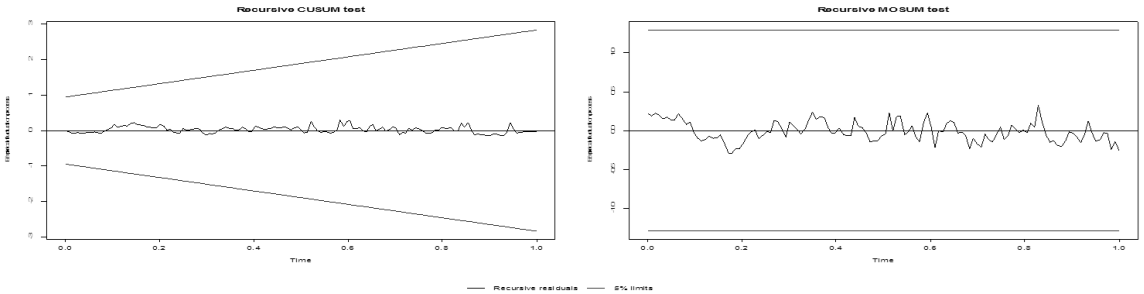

Figure 3. Plot of CUSUM and MOSUM: case of MP075.

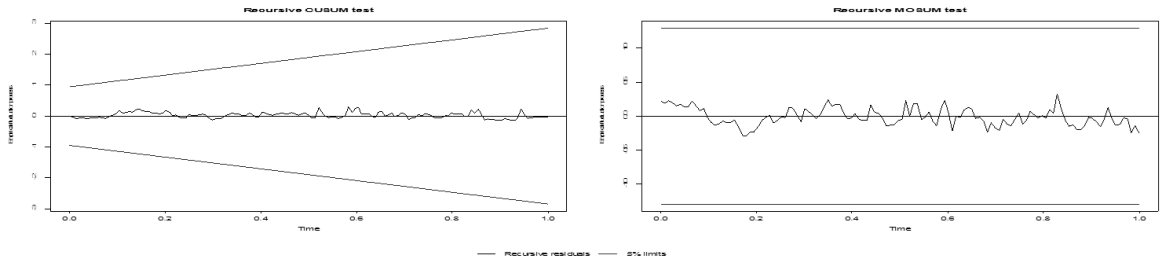

Figure 4. Plot of CUSUM and MOSUM: case of MP1.

REFERENCES

[1] Ali, S.B.; Selmi, S. and Hellali, W.: *Market power of Tunisian olive oil in EU market.* EuroMed Journal of Management **2**(3), 230-239, 2018, http://dx.doi.org/10.1504/EMJM.2018.10014465,

[2] Tasdogan, C.; Tsakiridou, E. and Mattas, K.: *Country market power in EU olive oil trade.* South-Eastern Europe Journal of Economics **2**(3), 211-219, 2005,

[3] Broos, E., et al.: *EU Wholesale Trade: Analysis of the Sector and Value Chains*. The Vienna Institute for International Economic Studies, wiiw Research Reports **415**, 2016,

[4] Ganapati, S.: *Growing Oligopolies, Prices, Output, and Productivity.* Center for Economic Studies, U.S. Census Bureau, Working Papers 18-48, 2018,

[5] Sheldon, I.M.: *The competitiveness of agricultural product and input markets: A review and synthesis of recent research.*
Journal of agricultural and applied economics **49**(1), 1-44, 2017,
http://dx.doi.org/10.1017/aae.2016.29,
[6] Filipović, F.: *Quality of foreign from the shelves and Croatian olive oil.* In Croatian.
http://www.niranaliza.hr/kvaliteta-stranog-s-polica-i-hrvatskog-maslinovog-ulja/blog,
[7] Suhartanto, D.; Dean, D.; Sosianika, A. and Suhaeni, T.: *Food souvenirs and their influence on tourist satisfaction and behavioural intentions.*
European Journal of Tourism Research **18,** 133-145, 2018,
http://dx.doi.org/10.54055/ejtr.v18i.317,
[8] Romo-Muñoz, R.A., et al.: *Heterogeneity and nonlinearity in consumers' preferences: An application to the olive oil shopping behavior in Chile*.
PLoS ONE **12**(9), 1-13, 2017,
[9] Pesaran, H.M. and Shin, Y.: *An autoregressive distributed lag modelling approach to cointegration analysis*.
In: Storm, S., ed.: *Econometrics and Economic Theory in the 20th Century*: *The Ragnar Frisch Centennial Symposium*. Econometric Society Monographs. Cambridge University Press, Cambridge, pp.371-413, 1999,
[10] Pesaran, M.H.; Shin, Y. and Smith, R.J.: *Bounds testing approaches to the analysis of level relationships.*
Journal of Applied Econometrics **16**(3), 289-326, 2001,
http://dx.doi.org/10.1002/jae.616,
[11] Toda, H.Y. and Yamamoto, T.: *Statistical inference in Vector Autoregressions with possibly integrated processes*.
Journal of Econometrics **66**(1-2), 225-250, 1995,
http://dx.doi.org/10.1016/0304-4076(94)01616-8,
[12] Goldberg, P.K. and Knetter, M.M.: *Goods Prices and Exchange Rates: What Have We Learned?*
Journal of Economic Literature **35**(3), 1243-1272, 1997,
[13] Folgado-Fernández, J.A.; Campón-Cerro, A.M. and Hernández-Mogollón, J.M.: *Potential of olive oil tourism in promoting local quality food products: A case study of the region of Extremadura, Spain.*
Heliyon **5**(10), 26-53, 2019,
[14] Hernández-Mogollón, J.M.; Di-Clemente, E.: Folgado-Fernández, J.A. and Campón-Cerro, A.M.: *Olive oil tourism: state of the art*.
Tourism and hospitality management **25**(1), 179-207, 2019,
http://dx.doi.org/10.20867/thm.25.1.5,
[15] López-Guzmán, T.; Cañero, P.; Moral-Cuadra, S. and Orgaz, F.: *An exploratory study of olive tourism consumers.*
Tourism and Hospitality Management **22**(1), 57-68, 2016,
http://dx.doi.org/10.20867/thm.22.1.1,
[16] Orgaz, F.; Moral, S.; López-Guzmán, T. and Cañero, P.: *Study of the existing demand in the field of oleotourism. Andalusia's house*. In Spanish.
Cuadernos de Turismo **39**, 437-453, 2017,
[17] Calzati, V. and De Salvo, P.: *The roll of gastronomic events in the promotion and valorization of rural areas. The house of Frantoi opened in Umbria.* In Italian.
FrancoAngeli, Milano, 2017,
[18] Moral-Cuadra, S.; López-Guzmán, T.; Orgáz, F. and Cañero, P.: *Motivation and satisfaction of the oleos of tourists in Spain. Andalusia's house.* In Spanish.
Espacios **38**(58), 4-17, 2017,
[19] Millán, M.G.; del Populo, M. and Sánchez-Rivas, J.: *Oleotourism as a Sustainable Product: An Analysis of Its Demand in the South of Spain (Andalusia).*
Sustainability **10**(101), 1-19, 2018,
http://dx.doi.org/10.3390/su10010101,

[20] D'Adamo, I.; Falcone, P.M.; Gastaldi, M. and Morone, P.: *A Social Analysis of the Olive Oil Sector: The Role of Family Business.*
Resources **8**(3), No. 151, 2019,
http://dx.doi.org/10.3390/resources8030151,
[21] Bimbo, F.; Roselli, L.; Carlucci, D. and de Gennaro, B.C.: *Consumer Misuse of Country-of-Origin Label: Insights from the Italian Extra-Virgin Olive Oil Market.*
Nutrients **12**(7), 21-50, 2020,
http://dx.doi.org/10.3390/nu12072150,
[22] Yan, J., et al: *Food fraud: Assessing fraud vulnerability in the extra virgin olive oil supply chain.*
Food Control **111**, 1-10, 2020,
[23] Ballco, P. and Gracia, A.: *Do market prices correspond with consumer demands? Combining market valuation and consumer utility for extra virgin olive oil quality attributes in a traditional producing country*.
Journal of Retailing and Consumer Services **53**, No. 101999, 2020,
http://dx.doi.org/10.1016/j.jretconser.2019.101999
[24] Hong, H. and Shum, M.: *Using Price Distributions to Estimate Search Costs.*
The RAND Journal of Economics **37**(2), 257-275, 2006,
http://dx.doi.org/10.1111/j.1756-2171.2006.tb00015.x,
[25] Hernández, J.M.; Folgado, J.A. and Campón, A.M.: *Oleotourism in the Sierra de Gata and Las Hurdes (Cáceres): Analysis of the potential of a test of a product.* In Spanish.
International Journal of Scientific Management and Tourism **1**(2), 333-354, 2016,
[26] López-Guzmán, T.; Cañero, P.; Moral, S. and Orgaz, F.: *An exploratory study of olive tourism consumers*.
Tourism and Hospitality Management **22**(1), 57-68, 2016,
http://dx.doi.org/10.20867/thm.22.1.1,
[27] Sánchez, J.D. and Ortega, A.: *The Olivarian Olive Cultivation: Historical Conformation, Patriotic Values and Cultural-Tourist Projection.* In Spanish.
Cuadernos de Turismo **37**, 377-402, 2016,
http://dx.doi.org/10.6018/turismo.37.256281,
[28] Shy, Oz: *Industrial Organization, Theory and Applications.*
The MIT Press, Cambridge, 1995,
[29] Vollrath, D.: *Fully Grown: Why a Stagnant Economy is a Sign of Success.*
University of Chicago Press, Chicago, 2020,
[30] Brown, R.L.; Durbin, J. and Evans, J.M.: *Techniques for testing the constancy of regression relationships over time.*
Journal of the Royal Statistical Society **37**(2), 149-192, 1975,
http://dx.doi.org/10.1111/j.2517-6161.1975.tb01532.x,
[31] Pesaran, M.H. and Pesaran, B.: *Working with Microfit 4.0: Interactive econometric analysis.*
Oxford University Press, Oxford, 1997,
[32] Zapata, H.O. and Rambaldi, A.N.: *Monto Carlo evidence on cointegration and causation.*
Oxford Bulletin of Economics and Statistics **59**(2), 285-298, 1997,
http://dx.doi.org/10.1111/1468-0084.00065,
[33] Wolde-Rufael, Y.: *Disaggregated industrial energy consumption and GDP: the case of Shanghai, 1952-1999.*
Energy Economics **26**(1), 69-75, 2004,
http://dx.doi.org/10.1016/S0140-9883(03)00032-X,
[34] Wolde-Rufael, Y.: *Energy Demand and Economic Growth*.
Journal of Policy Modeling **27**(8), 891-903, 2005,
http://dx.doi.org/10.1016/j.jpolmod.2005.06.003,
[35] Mavrotas, G. and Kelly, R.: *Old wine in new bottles: Testing causality between savings and growth.*
The Manchester School **69**, 97-105, 2001,
http://dx.doi.org/10.1111/1467-9957.69.s1.6,

[36] –: *Agricultural Market Information System*.
http://www.amis-outlook.org,
[37] Eurostat: *Nights spent at tourist accommodation establishments (tour_occ_nim).*
http://ec.europa.eu/eurostat,
[38] Sax C. and Steiner, P.: *Temporal Disaggregation of Time Series*.
The R Journal **5**(2), 80-87, 2013,
http://dx.doi.org/10.32614/RJ-2013-028,
[39] Narayan, P.K.: *The saving and investment nexus for China: evidence from cointegration tests 1979-1990*.
Applied Economics **37(**17), 2005,
http://dx.doi.org/10.1080/00036840500278103,
[40] Ang, J.B.: *Are saving and investment cointegrated? The case of Malaysia (1965-2003).*
Applied Economics **39**(17), 2167-2174, 2017,
[41] Elliott, G.; Rothenberg, T.J. and Stock, J.H.: *Efficient Tests for an Autoregressive Unit Root.*
Econometrica **64**(4), 813-836, 1996,
http://dx.doi.org/10.2307/2171846,
[42] Rebić, M. and Hoppe, I.: *Market information system in agriculture.* In Croatian.
http://www.tisup.mps.hr.

Interdisciplinary Description of Complex Systems 19(4), 542-560, 2021

TRUST, AUTOMATION BIAS AND AVERSION: ALGORITHMIC DECISION-MAKING IN THE CONTEXT OF CREDIT SCORING

Rita Gsenger[1,*] and Toma Strle[2]

[1]University of Vienna, Vienna University of Economics
Vienna, Austria

[2]University of Ljubljana
Ljubljana, Slovenia

DOI: 10.7906/indecs.19.4.7
Regular article

Received: 7 August 2021.
Accepted: 27 September 2021.

ABSTRACT

Algorithmic decision-making (ADM) systems increasingly take on crucial roles in our technology-driven society, making decisions, for instance, concerning employment, education, finances, and public services. This article aims to identify peoples' attitudes towards ADM systems and ensuing behaviours when dealing with ADM systems as identified in the literature and in relation to credit scoring. After briefly discussing main characteristics and types of ADM systems, we first consider trust, automation bias, automation complacency and algorithmic aversion as attitudes towards ADM systems. These factors result in various behaviours by users, operators, and managers. Second, we consider how these factors could affect attitudes towards and use of ADM systems within the context of credit scoring. Third, we describe some possible strategies to reduce aversion, bias, and complacency, and consider several ways in which trust could be increased in the context of credit scoring. Importantly, although many advantages in applying ADM systems to complex choice problems can be identified, using ADM systems should be approached with care – e.g., the models ADM systems are based on are sometimes flawed, the data they gather to support or make decisions are easily biased, and the motives for their use unreflected upon or unethical.

KEY WORDS

algorithmic decision-making, credit scoring, trust, automation bias, algorithmic aversion

CLASSIFICATION

APA: 2910

JEL: O3

*Corresponding author, η: rita.gsenger@wu.ac.at; +436 991 072 6227;
Heinrich-Collin-Str. 8-14/6/24, 1140 Vienna, Austria

INTRODUCTION

The process of decision-making is prone to a diverse range of biases that can lead, at least in certain contexts, to erroneous judgments or disadvantageous choices (e.g., [1-4]). From the perspective of classical (especially economic) models of decision-making – where the decision-maker is seen as a kind of a globally rational agent (e.g., [5]), approximately capable of and motivated to maximise her utility – the range of contexts where people systematically fall prey to erroneous judgment or make disadvantageous choice is remarkable. For instance: people are quite prone to irrelevant anchors when making judgments or choices [4, 6]; much more responsive to losses than gains [7]; people's choices are strongly affected by how choice problems are formulated [8, 9]; people are likely to choose smaller, immediate rather than larger, more distant rewards [10]; prone to remain with default choices even if disadvantageous [11] and quite indecisive [12]; people even choose disadvantageously from the perspective of their own happiness [13]; and are even blind to reasons of their own, seemingly deliberate and simple, choices [14]. Decision-makers are, in a rather important sense, quite bounded in their "rationality". As Herbert Simon lucidly states: "[…] the concept of "economic man" (and, I might add, of his brother "administrative man") is in need of fairly drastic revision … Broadly stated, the task is to replace the global rationality of economic man with a kind of rational behavior that is compatible with the access to information and the computational capacities that are actually possessed by organisms, including man, in the kinds of environments in which such organisms exist." [5; p.99].

Several ways of amending the imperfection of human decision-making are available. One strategy is to try to educate decision-makers through various debiasing strategies (e.g. [15]). Another is to modify choice environments and thus help decision-makers make better decisions – e.g., the strategy of the nudge programme [11]. Another solution, increasingly used in administrative and economic sectors, is to use algorithms to support human decision-making or to entrust decision-making to algorithmic systems altogether.

Some would say that using algorithms for decision-making purposes promises a reduction of biases in judgment and decision-making as they are, for instance, able to consider more information [16]. Furthermore, some have argued that ADM systems enable fairer and more objective decisions, since algorithms are not, for instance, affected by emotions [17], or because their decision-making process is, at least in principle, more transparent and accountable than humans' [18]. Moreover, ADM can provide relief from cognitive workload of users and decision-makers having to make choices in a rather complex world [19-21]. ADM systems have been, for these and other reasons, employed in many different contexts, such as to determine loans [16, 22] and insurance premiums [23], to investigate tax evasion [24], to calculate credit scores [25-27], to predict the likelihood of criminal activity [16, 28-30], in policing [31-33], healthcare [34, 35] and within social media platforms [36, 37].

All in all, the ubiquitous use of ADM systems has widespread economic and social consequences, as it is transforming business sectors and creating new ways of social organisation [16]. It must be noted, however, that the consequences of using such systems can be quite dire: from biased models leading to disadvantaging the already marginalized groups to enabling new and effective ways of manipulating people's behaviour [16]. The consequences, risks, and ethical questions within ADM should thus be taken seriously and critically reflected upon [16, 18, 38, 39]. Considering the whole range of consequences and risks involved in using ADM systems surpasses the scope of this article.

Instead, we aim to understand how people's attitudes towards and beliefs about ADM systems affect their use, influence, and success of application. In the first part of the article, we briefly discuss some main characteristics and types of ADM systems. In the second, we introduce the functioning and employment of ADM systems and argue that their results are

often perceived differently from human recommendations. In the third part, we focus on trust, automation bias, complacency, aversion, and the resulting behaviours, drawing on research from various areas such as psychology, human-computer-interaction, and cognitive science. In the fourth part, we apply findings from previous research to the context of credit scoring where such research is scarce. Credit scoring is an interesting use case for ADM systems primarily due to its application in various domains having a pervasive influence on many areas of people's lives. Moreover, in credit scoring customers can hardly object to a credit score calculated by an ADM system. To affect their credit scores certain groups of people, for instance, engage in "strategic data-generating performances" [25, p.349], deliberately creating data to produce a more favourable credit score [25]. We end the article with a brief discussion of ADM systems for credit scoring from the perspective of human-centric approach to AI and touch upon certain ethical issues and challenges.

ALGORITHMIC DECISION-MAKING SYSTEMS

Various kinds of automated or partially automated systems that support decision-making can be distinguished. Castelluccia and Le Métayer [18] suggest that three classes of systems can be distinguished. First, systems that aim to improve knowledge and technology by analysing big datasets to support, for example, climate forecasts or research in healthcare (e.g., to assist the process of discovering a new virus). Second, systems that support decisions by making recommendations and predictions, utilised, for instance, "to improve logistics (optimal product placement in stores, optimal road constructions or the frequency of refuse collection), finance (real-time auctions) or security (automated detection of vulnerabilities in computer systems)" [18; p.5]. Systems of that category are used additionally to optimize and improve services that have been performed by humans so far. Third, systems that enable inanimate objects to act and decide to some extent on their own. In this context, the algorithms, for example, autonomous cars or robots, are making decisions on behalf of the users. (Anthropomorphic systems such as robots were excluded from our analysis as these might elicit different reactions due to their form [40].)

Some ADM systems allow operators or users control over recommendations, suggestions, and the degree of the systems' use. For instance, in social media, users can disable suggestions about personalised advertisements or content [36, 37]. In the context of managerial or governmental decisions, however, people often have little choice in following recommendations and using these systems [20]. Generally, different levels of automation can be distinguished: from no assistance by the computer to full automation whereby the ADM system decides and/or acts autonomously. The computer could, for instance, provide multiple or only one option, wait for the approval or allow a veto by a human operator, provide information only if asked or solely if the computer decides to do so. Higher automation levels might be beneficial for tasks that do not require flexibility and systems with a low chance of failure [40].

Here, we will be concerned with ADM systems giving recommendations to support the decision-making process of a user or operator, irrespective of their influence on the decision outcome. We will consider systems that perform services or parts of services that humans used to be in charge of, such as credit scoring. As we will focus on various attitudes towards ADM systems and the resulting behaviours of users or operators, we will consider different kinds of individuals influenced by or related to ADM systems: operators and users (like customers of a bank), developers and designers of the systems, etc.

ATTITUDES TOWARDS ALGORITHMIC DECISION-MAKING SYSTEMS

Perceptions of and attitudes towards ADM systems can be investigated on several different levels. First, the perception varies between stakeholders and people, such as designers and

developers, operators, users, the media and the public. Second, the attitude toward the systems might affect the perceptions of their decisions [20]. Therefore, the latter needs to be considered as it influences the successful employment of such systems. Third, the perception can vary according to the influence of individual aspects, for instance, the cultural background of the users [41] or their expertise and knowledge [19, 42]. Overconfidence in the algorithmic systems might lead, for instance, to their unnecessary use [43]. Furthermore, peers and/or the media might alter certain expectations people have about ADM systems. Moreover, they cause a different perception of the advice given by ADM systems compared to advice from humans, even if the content of the advice itself is the same [42]. Algorithmic systems might be preferable to human decision-makers in some contexts as they outperform human experts in prediction across different domains such as climate forecasts [44], the discovery of new viruses [18], and clinical diagnosis [45].

According to Lee and See [46], the perception of and the beliefs about ADM systems might be positive due to various reasons: First, users and operators might judge them more apt and objective [47, 48], as more information is readily available to them [28, 46]. Second, the systems are less influenced by emotions. Therefore, their decision-making might be more competent [20]. Third, in some instances, users and operators perceive systems as value-neutral in their decision-making [28]. However, for some choices, intuition is understood as useful or even required – accordingly, users might perceive systems as less competent in decision-making [49].

TRUST IN ALGORITHMIC DECISION-MAKING SYSTEMS

Algorithms can only be useful to support human decision-making if users, operators, and stakeholders trust them [50]. Gaining trust is influenced by expectations [51], familiarity [52, 53] and non-verbal cues during an interaction [54]. In psychological, behavioural, and neuroscientific research, trust has been described as an attitude [46], a behaviour [55], a relationship [56], and a brain activation pattern [57]. Trust in any automated system includes specific influences such as reliability, utility, robustness, and a false-alarm rate [58]. Moreover, research has shown that some people tend to trust automated systems more and perceive them as more reliable than human individuals, a phenomenon called the automation bias [59] (for details on the automation bias, see section 2.3). Overall, trust in automated systems depends largely on performance, such as the response time of the system [60]. The process of establishing trust depends on the operator's knowledge about the system, its design features, and other situational influences such as the expertise of the truster [41].

The participants of a study about the trustworthiness of ADM systems [20] regarded both humans' and algorithms' decisions as equally trustworthy if they concerned scenarios of mechanical tasks. Algorithmic decisions were perceived as less trustworthy in more human tasks, such as scheduling in the workplace. Most participants were aware that an algorithmic system could exhibit glitches. Therefore, no participant trusted the algorithm without reservations [20].

Adopting trust towards automated systems facilitates the navigation of complexity, replaces supervision, and enables reliance when the system is too complex to be understood completely. Reliance, however, cannot always be accurate in terms of the capabilities of the automated system. Blind reliance on automated systems can be just as detrimental to its application as not trusting the system at all. If operators trust the system blindly, mistakes might not be detected. If they do not trust it at all, cooperative decision-making is not possible. Trusting the system too much or not enough might be described in terms of misuse and disuse: "Misuse refers to the failures that occur when people inadvertently violate critical assumptions and rely on automation inappropriately, whereas disuse signifies failures that

occur when people reject the capabilities of automation" [46; p.50]. Inappropriate reliance resulting in disuse and misuse of automation is frequently caused by a mismatch between the system's capabilities and the trust invested. This discrepancy is described in terms of (i) calibration, (ii) overtrust, and (iii) resolution [46]. The first aspect, calibration, refers to a mismatch between the trust invested and the system's capabilities. Overtrust concerns the phenomenon of trusting the system too much due to poor calibration. Resolution describes "how precisely a judgment of trust differentiates levels of automation capability" [46; p.55]. If the resolution is low, large changes in the system are met with small variations of trust. Misuse and disuse of automated systems can be decreased by greater specificity – meaning the flexible adaptation of trust over time, high resolution, and good calibration of trust in the system's capabilities [46].

Estimating a system's capabilities correctly and placing enough trust in it depends on the knowledge about its capabilities and functioning, as a study by Alexander et al. [19] on over – or underreliance on ADM systems has shown. In the study, participants were given recommendations by ADM systems in a problem-solving game. To make the choice nontrivial, participants had to pay the algorithm to support them in making money. Participants had to solve two-dimensional mazes, getting a reward of 5 $ if they solved one in 60 seconds or less. The support of the algorithms would cost 2 $ and they could either adopt the suggestions of the algorithm or ignore them. Participants were in one of four conditions with varying information about the system: the first group was not given any information about the suggested algorithm; the second one was told the algorithm had a 75 % accuracy rate; the third group was told that 54 % of people used this algorithm; the fourth group was told that 70 % of people used it. The study measured the neurophysiological response, cardiac rate, and behaviour of participants to determine if they relied too much or not enough on the algorithm. By measuring heart rate variability, researchers determined the cognitive load and arousal of participants. According to the study, the social proof was the most effective tool in convincing people of the algorithmic system. Moreover, the study found that the adoption of the algorithm reduced cognitive load in all conditions. That might suggest that participants did not monitor the algorithm after its adoption. Therefore, the performance was lowest when participants included the algorithm. Generally, the attention of participants adopting the algorithm was lower than non-adopters', but it was still higher than baseline, meaning they did pay attention to what the algorithm was doing.

AUTOMATION COMPLACENCY

In supporting human decision-makers, ADM systems are often designed to reduce erroneous judgments. However, they can cause other types of errors, such as automation complacency, leading to disadvantageous decisions. Automation complacency is defined as a human operator monitoring an automated system and missing a system failure or malfunction due to substandard monitoring [21]. Complacency as well as automation bias (for details see the next section 2.3.) were first researched in the aviation sector [61]. There, pilots, air traffic controllers, and other responsible personnel, can underestimate threats and work under the assumption that everything is fine, even though there is evidence to the contrary. Their negligence ultimately results in an accident. The term automation complacency was coined regarding automated aviation systems [21]. Operators of automated systems mostly passively observe and control the functioning of the system. Even as that has increased speed and efficiency, automation also gave rise to the misuse of automation [62]. Due to the assumption that everything is working correctly, operators insufficiently inspect automated systems compared to systems under manual control. Consequently, system malfunction or failure might be missed, or reactions might be delayed [62].

Parasuraman and Manzey [21] have shown that complacency occurs especially for highly reliable systems. The detection rate of failures increases if the system is not entirely reliable. However, these results vary depending on the expertise of the participants. Accordingly, Parasuraman and Manzey [21] distinguish between complacency potential and behaviour, whereby the latter occurs only if the potential is given with other circumstances, such as a high workload [21]. Moreover, research indicates that complacency might be a compensatory mechanism to deal with a high workload [62].

Complacency in ADM systems has been researched in the context of the control problem. The problem arises when operators supervise a task execution, which is increasingly the case, for instance, in aviation, where the plane flies automatically while the pilot monitors the situation. When using reliable automated systems, pilots might become "complacent, overreliant or unduly diffident when faced with the outputs" [63; p.556]. The complacency effect affects novices as well as experts and might have damaging consequences such as accidents. Generally, the less human intervention is necessary for a system to function, the greater the likelihood of the control problem occurring [63].

More recently, automation complacency has been observed in decision-making systems based on machine learning. For instance, in predictive policing, officers go on the recommended route without challenging it. Otherwise, they would have to justify the divergence from a fixed procedure dictated by the algorithm [64]. In another example, as shown by Eubanks [38], caseworkers who are responsible for child welfare in Pennsylvania and work in a governmental agency using an ADM system were more inclined to adapt their own risk estimates to the model's estimates instead of taking advantage of the scope of action they had [38].

AUTOMATION BIAS

Automation bias refers to the human tendency of ignoring or not inquiring about contradictory information about a computer-generated solution, which is accepted as correct [61]. Moreover, automation bias is enforced when a system gives the wrong advice, whereas complacency occurs if the system does not give advice, even though it should [65]. Automation bias is defined, similarly to other biases, as the use of a "heuristic replacement for vigilant information seeking and processing" [21; p.391], but contrary to other decision biases, it results specifically from the interaction with an automated system [21].

Parasuraman and Manzey [21] identify three causes of automation bias: 1) The cognitive-miser hypothesis, stating that humans prefer to reduce their cognitive load and thus decide according to simple decision rules and comprehensive heuristics, which might result in automation bias, as operators do not undertake any thorough analysis; 2) Automated systems are perceived as powerful agents, believed to have more analytic capabilities than humans, and thus they are trusted more; 3) Responsibility might be handed over to the automated system, as people try to reduce their own contributions when work is shared. If the automated system is part of a team, other team members might reduce their efforts and refrain from analysing additional aspects or inspecting the automated system's decisions. Studies suggest, however, that systems that support analysis and information integration are less prone to lead to automation biases than systems, which recommend specific actions due to their analysis [21]. Therefore, automation bias might occur due to cognitive overload and could be reduced by decreasing cognitive load [35]. According to Parasuraman and Manzey [21], the effects of automation bias are twofold: First, operators following incorrect recommendations are committing a commission error. Second, an error of omission happens when operators neglect a critical situation because they were not informed by the system (see, for instance, the Enbridge Pipeline Disaster [66]).

Reducing cognitive load, trusting algorithmic systems more than humans, and handing over responsibility increase the occurrence automation bias. Moreover, the degree to which operators perceive themselves socially accountable – for instance, when in direct interaction with a customer to whom they must justify a decision – plays a role regarding the frequency of omission and commission errors. People who feel accountable were more thoroughly examining the decisions taken by an algorithmic system and verifying them more often [21, 67]. Accountability creates pressure for people to include more information and process it more thoroughly. Moreover, accountability enables decision-makers to "employ more multi-dimensional, self-critical and complex information processing strategies and to put more effort into identifying appropriate responses" [67; p.703].

ALGORITHMIC AVERSION

The previous sections highlighted the misuse of automated systems due to overestimation, complacency, and bias. Here we describe the phenomenon of algorithmic aversion where a negative attitude towards ADM systems might lead to a disregard of their help compared to a person's advice [68]. As Dietvorst et al. spell out various characteristics of algorithmic aversion: people often "prefer humans' forecasts to algorithms' forecasts, […] more strongly weigh human input than algorithmic input, […] and more harshly judge professionals who seek out advice from an algorithm rather than from a human" [69; p.114].

Additionally, the type of the decision plays a role regarding the degree of aversion towards an ADM system. Lee [20] has shown that participants have similar emotional responses to decisions made by algorithms and humans if the decision requires solely mechanical skills and does not require subjective judgment or emotions. Conversely, regarding human skills, emotions towards the systems' decisions were more negative compared to humans' decisions. Generally, participants felt less positive about managerial decisions made by algorithmic systems [20].

After giving bad advice, advice utilisation decreases more for an automated system than for a human advice giver. This probably happens because people expect automated systems to be more "perfect" compared to "flawed" human beings. Moreover, participants might have confidence in human advisors to perceive and correct their own errors [70]. People were prone to prefer the advice of ADM systems compared to humans before mistakes in their decisions were known to them [69, 70]. This seems to be consistent with multiple studies done by Logg et al. [71] where people clearly preferred the recommendations by ADM systems (regarding forecasts of popular songs, romantic attraction, and numeric estimates).

Expectations about the functioning and the capabilities of such systems shape the users' perceptions [42]. Multiple studies done by Yeomans et al. [72] reveal a considerable aversion towards ADM systems if the functioning of the systems is not known. Participants preferred human recommendations as they reportedly understood them better. Therefore, the researchers conclude that not only increased accuracy, as often suggested, but also an understanding of ADM systems would decrease aversion. Overall, the aversion against these systems depends to some degree on the understanding of and knowledge about those systems [72]. These studies contradict the findings of Logg et al. [71] where experts on ADM systems and forecasts tended to rely less on the recommendations by ADM systems than ley people.

The degree of aversion depends on the expectations and beliefs about the algorithm influence, the need for control by users, and the capability of alignment with the outcome of an algorithmic decision. Often, ADM systems dominate the decision-making instead of enabling a transparent process that includes the algorithm and the human in an aligned decision-making process. Algorithmic aversion develops if the consequences following a decision are not the same as expected, and the human user loses confidence in the system [42].

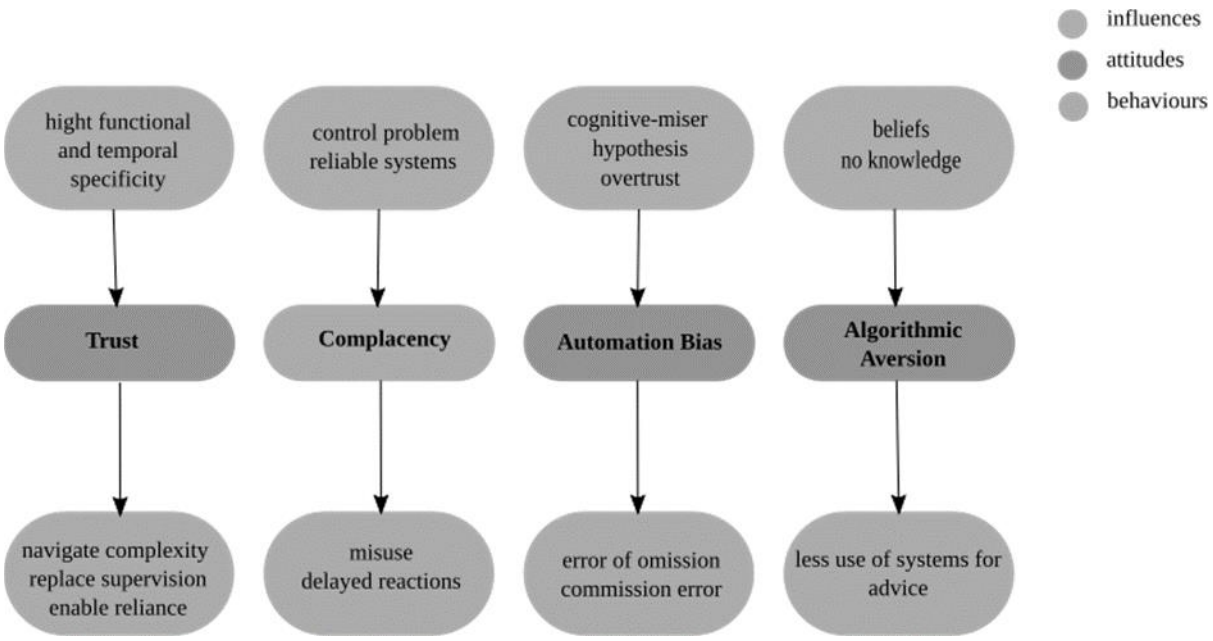

Figure 1. Summary of influences on and effects of trust, complacency, automation bias and algorithmic aversion.

CREDIT SCORING

In this part, the functioning and the use of credit scoring is first briefly explained. Second, since insight into the attitudes towards ADM systems within credit scoring is limited, we consult research and theory on attitudes towards ADM systems from other domains (presented in the previous sections) and apply it to credit scoring.

WHAT IS CREDIT SCORING?

Traditionally, if the customer of a bank wanted to get a loan, she would need to go to the bank and be interviewed. Subsequently, a credit manager would evaluate her trustworthiness, reliability, and the risk of defaulting [16]. Furthermore, the evaluation would rely on her financial status and factors such as marital status, gender, address, employment, housing and criminal history.

However, the evaluation system has been increasingly computerized since the 1950ies, leading to centralization and standardization of the evaluation process. Due to computerization, requests can be processed faster, and unskilled workers are hired instead of skilled bankers, reducing personnel costs [26]. Moreover, the evaluation systems seem to have removed the human element from credit scoring by excluding personal contact between borrower and lender. By doing so, the systems promised to remove the prejudice credit managers might have towards customers. Additionally, they "reduced personal creditworthiness to the sum of statistical probabilities" [26; p.232]. An applicant is perceived as a part of a risk population, determined by demographic and economic qualities. Therein, intervention of credit managers is viewed as a distortion. Moreover, the process of determining creditworthiness is seemingly separated from characteristics such as honesty, responsibility, and morality of the customer [26]. Recently and increasingly, companies are trying to improve the credit scoring system by evaluating creditworthiness based on personality traits such as patience, impulsiveness, risk preference, and trustworthiness, among others [73], including social network data [27].

Credit scoring is used worldwide in different domains such as individual loans, mortgages [74] or contracts such as telephone contracts [75]. The mechanism applies a statistical model "that tries to predict the future behaviour of accounts and customers based on data from the same or a similar group of accounts and customers from the relatively recent past" [74; p.59]. Every business uses its own model based on different methodologies. In the US, the Fair Isaac Corporation (FICO) developed an algorithm used by three major credit reporting agencies. The algorithm itself is unknown but it is most likely based on the ratio of debt to available credit [73]. Dozens of different commercially available scores for different kinds of debts can be distinguished, such as credit card debt, personal loans, or automobile loans [26]. A model produces a certain score, wherein a low score would mean low quality and high risk. The acceptance of a credit application and the credit conditions are decided according to the determined score. Edelman, however, points out that credit scoring is additionally a business process, which includes the "data quality, credit policy, profitability, model stability and what to do with decisions that the bank or the branches think are not correct" [74; p.60]. Until the late 1980s, only data from the credit application informed the decision of the credit scoring system, meaning there was one evaluation to determine creditworthiness. Nowadays, however, the creditworthiness is evaluated continuously after the acceptance of a credit application. These evaluations are done by scoring algorithms that examine if customers pay in time and if they are still profitable for the credit bureau, developing risk models. Risk is low enough if they are "carrying an interest-generating balance without maxing out" [26; p.252], meaning defaulting on the credit. Simultaneously, the risk and the performance of an individual can be tracked across all her accounts, even if she borrowed money from multiple credit bureaus [26]. Scorecards play an important role in determining creditworthiness. These are tools embedded in software packages to select customers and calculate credit scores. The software includes "back-stage statisticians, electronic data warehouses, risk managers, and front-stage marketing campaigns" [76; p.284]. Scorecards differ according to the way they translate the conditions in which risk is analysed.

Calculating credit scores by using statistical models can operate under more certainty if data about the consumers are available, which do not stem from the consumers themselves. Digital data analysis allows for replacing dependency of information given by the consumers directly to the credit lending institution [76]. The data used for statistical credit scoring comprises up to 400 variables provided by credit reference agencies [74]. In the era of big data, often no differentiation between credit data and other data is made and many other variables which are not connected to the credit history of a customer are included [27, 77-80]. For instance, an applicant's college, her use of capital letters in applications (whereby the use of all caps writing is interestingly a warning sign) and social media data [26], including online tracking and behavioural profiling [79, 81]. Moreover, data harvested by specific apps from smartphones might be included [82]. Furthermore, other network-based data is included, developing a social credit score based on the individuals' position in a social structure [79]. Additionally, the inclusion of network data allows targeted advertising of credit products to new customers [83] and the inclusion of individuals who previously did not have access to credit [77].

As for the previously used consumer credit assessment, which only included data from the application made by the customers concerning their credit history, three factors are generally assessed through credit scoring: Stability, honesty, and the ability to repay a credit. This assessment is usually repeated on a regular basis [74]. Paying back debt depends not only on the ability to do so but also on the willingness to pay. Behavioural tendencies, for instance, trustworthiness, reliability, impulsivity, and risk attitude are used as defining characteristics to determine individuals' willingness to pay back the debt [73]. For a detailed history of credit assessment and the quantification of creditworthiness, see [26]. For a historical perspective on the FICO score, see [76].

Credit scores determine evaluations of other areas of people's lives as well. For instance, some employers use the credit score to determine if the customer is a responsible employee and trustworthy [73]. Individuals who pay their bills on time are presumed to be responsible in the workplace as well, not accounting for many other factors that could cause a bad credit score. That might lead to a negative feedback loop as people with bad credit subsequently have more difficulties finding a job, making their credit even worse [16]. Moreover, sometimes tax inspectors use credit scores to decide whom to investigate [74].

ATTITUDES TOWARDS CREDIT SCORING ALGORITHMS

Different attitudes and behaviours are formed while using ADM systems that we have described in chapter two: trust, complacency, automation bias, and algorithmic aversion. Each of these attitudes has different causes, dependencies, and results (see Figure 1). Some of them can be observed or applied to the use of ADM systems in credit scoring.

In the context of credit scoring, systems barely permit human intervention. Often, employees are required to use the system to calculate the credit score and the contract's conditions [26]. Moreover, not being able to influence the credit score even if it is perceived as unjust might increase algorithmic aversion.

Furthermore, credit decisions are not mechanical, making them possibly inept to be taken over by such systems. Algorithms used for credit scoring are criticized for introducing standardization in a highly complex area and entailing negative consequences for individuals [16].

Burton et al. [42] suggest that ADM systems as support along every step of a decision-making process would enable their adoption by more users and include more application domains. Such differentiated systems might be beneficial for credit scoring as well, as increasing flexibility and adaptation to the needs of customers. Burton et al. [42] define the successful use of ADM systems as the shared decision-making capability of the human operator and the system. Complete trust in or disregard of the systems point to the failure of the interaction between the operator and the system, even as in some application domains, full automation is beneficial [42]. That, however, does not seem to be the case for ADM systems used in credit scoring, as a shared agency, and avoidance of complacency and aversion could benefit the person affected by a credit score. Combining the capacities of the systems, considering large quantities of data on the one hand and the knowledge of human circumstances and individual situations, on the other hand, could make such systems more successful. Overall, a more human-centred approach to ADM systems would be beneficial for their successful use, as such an approach could solve problems of accuracy, bias, and transparency [84].

A study done by Schäufele [75] shows that operators of credit score systems often do not question the decisions of the system even if they could, indicating automation bias. Operators in that study did not see any reason to question the system even though they could object to decisions by filing a complaint. They seem to give away responsibility for the decision to the ADM system. Moreover, operators reduce their cognitive load by relying on the system too much for the complex credit score calculation [75]. These factors indicate automation bias in credit scoring systems.

As previously mentioned, automation bias might lead to commission and omission errors [21]. In the case of credit scoring, commission errors seem to occur, as the type of contract a person gets depends on the inference of data from a profile made about her, using a model that might be biased. The commission error has dire consequences for some, who receive, for instance, bad conditions for a credit, which keeps them in debt. These consequences are especially difficult as a bad credit score influences other domains such as employment. In

consequence, people have more difficulty paying back their debt as they are not hired. Automation bias thus results in a negative feedback loop [16].

Furthermore, uncritically accepting credit scores could let customers believe the systems' recommendations without comparison to other credit providers [62], thus causing complacency [16].

All in all, credit scoring would need to account for the unique and complex life situations of very different individuals and situations. And although ADM systems promise to reduce biases of decisions about credit scores, and the systems have more data available and a bigger processing capacity, the models employed are still human-made, mostly not user-centred, and thus criticized as biased [16], perceived to be unfair to certain individuals and can lead to negative feedback loops. For instance, some groups of people, to circumvent a negative credit score, even "play [...] the credit score game" [25; p.346], finding strategies to improve their credit score (also to enable upward social mobility [78]) by producing positive data; for instance, by joining lending circles where people lend money to each other without interest to build credit [85] (see also [86-88] for similarly created loops within systems, rich with social interaction).

HUMAN-CENTRIC ADM SYSTEMS FOR CREDIT SCORING

The attitude towards ADM systems is crucial for their successful and beneficial use. A survey conducted among U.S. adults in 2018 by the Pew Research Center shows that 31 % of participants deem using automated decision-making systems for credit scoring acceptable for companies. Respondents who would not consider such a system acceptable voiced concerns such as the violation of privacy, the accuracy of online data representing a person, and the irrelevance of online habits and behaviours for an individual's creditworthiness [89]. The exploration and increasing use of alternative data sources for credit scoring, including social media data or information from smartphones [79, 82, 83, 90], is perceived rather critically in research [27, 83] and by most participants as inquired by the Pew Research Center [89]. Users and customers (people who these systems decide about) seem to have a different attitude towards ADM systems than managers and executives of companies or institutions who decide about these systems. The latter seem to emphasize the timeliness and efficacy of their companies due to using these systems [28]. To alleviate users' concerns and to make ADM systems successful and possibly fairer, a human-centric framework is necessary.

A human-centred framework provides strategies to use AI systems to improve capabilities instead of replacing the workforce, including "*human factors design* to ensure AI solutions are explainable, comprehensible, useful, and usable" [91; p.44] (emphasis in original). Ethical design principles to ensure fairness and justice [91] are crucial for an AI systems' human-centric framework. The social accountability of operators and managers is important to establish and maintain the fairness of systems. Studies show that participants who know to be accountable when using ADM systems committed significantly fewer omission and commission errors than control groups [21]. Another study by Lee and Baykal [92] found that interpersonal power as well as knowledge of programming influence the attitude towards decisions by mathematically fair algorithms compared to group decisions. (Fair division algorithms use a mathematical definition of fairness, which in most cases uses equity as a central concept. Equity, in contrast to equality, does not advocate for treating every person the same, but accounts for individual differences to guarantee a fair distribution [92].) Their results show that participants perceived decisions made through group discussions as fairer. Furthermore, algorithmic decisions were thought to be unfair if the algorithm did not "account for multiple concepts of fairness and cognitive and social behaviours in groups, such as the presence of altruism and group dynamics" [92; p.1035]. The authors attribute the

increased perception of fairness in group discussions to the decision's transparency and the possibility of individual group members' intervention. As individuals were held accountable, the perception of fairness increased [92]. Overall, adapting the decision-making algorithm to specific tasks by increasing functional and temporal specificity might ensure fairness and reduce algorithmic aversion. Moreover, the social accountability of the operators could reduce omission and commission errors, especially for group decisions.

Aside from social accountability, a legal framework is necessary to regulate the use of ADM systems and ensure algorithmic accountability. The General Data Protection Regulation 2016/679 (GDPR) grants several new rights to citizens of member states of the European Union, including the right to be forgotten and to have their data deleted (Art. 17) or rectified (Art. 16), the right to be informed about ADM systems and their use, including the consequences of such systems and profiling (Art. 13) and the right not to be subject of an automated decision, which includes profiling (Art. 22). Exceptions to Article 22 of the GDPR can be granted if (1) the data subject's informed consent is provided, (2) if the legislation of a member state allows such decision-making, or (3) if the decision is necessary to fulfil a contract (Art. 22(2)). Furthermore, the GDPR grants the right to a human reassessment of the system's decision if perceived as unfair or incorrect [93].

Critics claim that the legal framework provides too much freedom to data controllers and insufficiently protects individuals [94]. Furthermore, as ADM systems are very complex, the information should be presented in a comprehensible manner for each individual, and the system's "intentions" made clear [94]. What is problematic is that, not all parties involved have a right to an explanation, for instance, the general public. Providing the public with information concerning the ADM systems' functioning, however, would be beneficial to reduce public concerns and to improve individuals' understanding of the use of their data [94]. By increasing knowledge about these systems, social accountability could be created, and influence could be exercised to make these systems more human-centric. Especially regarding credit-scoring systems, which possess sensitive data about individuals, creating accountable and transparent systems is crucial to ensure a fair distribution of credit.

CONCLUSION

This article aimed to identify peoples' attitudes towards ADM systems and ensuing behaviours when dealing with ADM systems with a particular consideration for credit scoring.

Trust and algorithmic aversion were identified as common attitudes adopted towards ADM systems, automation bias and complacency as key behaviours. Trust, complacency, automation bias, and algorithmic aversion were consulted to shed light on the attitudes towards ADM systems for credit scoring. In credit scoring, all these aspects could be identified, complacency resulting in overreliance and automation bias engendering commission errors by the operators, causing the misuse of ADM systems. Moreover, complacent users might not question the credit score assigned to them. ADM systems' decisions could be most beneficial for the service providers because they might be designed to find the most cost-efficient solutions leading to complacency by the operators and managers. These solutions might not be the most beneficial for the users or customers who must live with the consequences.

Furthermore, aversion could be influential for operators and users. First, credit scoring systems do not allow for human interference, and second, they might not be perceived as fair. Multiple strategies are suggested in research to reduce errors and biases, such as highly functional and temporal specificity and a human in the loop, to reduce complacency effects [46]. Moreover, social accountability and transparency of the decision-making process by the algorithmic

systems might be useful strategies to establish trust on the one hand and reduce bias, aversion, and complacency on the other hand. The design of human-centred ADM systems would benefit customers and operators alike. That, however, would require designing systems based on explainability and transparency instead of data that are often opaque and biased but are used due to easy access and availability [89].

All in all, ADM systems are increasingly used, influencing decisions made by companies, policymakers, and individuals [18]. Even as these systems are frequently advertised to be more objective and reliable than human decision-makers [28, 46], and many advantages in applying ADM systems to complex choice problems can be identified, using ADM systems should be approached with care since they are sometimes based on biased models, and the motives for their use unreflected upon or unethical. Often unreflected use of ADM systems might thus too easily result in dire consequences for individuals and, more often than not, for the already disadvantaged groups [16, 38, 39].

REFERENCES

[1] Kahneman, D.: *A perspective on judgment and choice: Mapping bounded rationality*. American Psychologist **58**(9), 697-720, 2003, http://dx.doi.org/10.1037/0003-066X.58.9.697,

[2] Kahneman, D. and Klein, G.: *Conditions for intuitive expertise: A failure to disagree*. American Psychologist **64**(6), 515-526, 2009, http://dx.doi.org/10.1037/a0016755,

[3] Thaler, R.H. and Sunstein, C.R.: *Nudge: Improving Decisions about Health, Wealth and Happiness*. Yale University Press, London, 2008,

[4] Tversky, A. and Kahneman, D.: *Judgment under Uncertainty: Heuristics and Biases*. Science **185**(4157), 1124-1131, 1974, http://dx.doi.org/10.1126/science.185.4157.1124,

[5] Simon, H.A.: *A Behavioral Model of Rational Choice*. The Quarterly Journal of Economics **69**(1), 99-118, 1955, http://dx.doi.org/10.2307/1884852,

[6] Englich, B.; Mussweiler, T. and Strack, F.: *Playing Dice With Criminal Sentences: The Influence of Irrelevant Anchors on Experts' Judicial Decision Making*. Personality and Social Psychology Bulletin **32**(2), 188-200, 2006, http://dx.doi.org/10.1177/0146167205282152,

[7] Kahneman, D. and Tversky, A.: *Prospect Theory: An Analysis of Decision under Risk*. Econometrica **47**(2), 263-291, 1979, http://dx.doi.org/10.2307/1914185,

[8] Ruggeri, K., et al.: *Replicating patterns of prospect theory for decision under risk*. Nature Human Behaviour **4**, 622-633, 2020, http://dx.doi.org/10.1038/s41562-020-0886-x,

[9] Tversky, A. and Kahneman, D.: *The Framing of Decisions and the Psychology of Choice*. Science **211**(4481), 453-458, 1981, http://dx.doi.org/10.1126/science.7455683,

[10] Green, L.; Fry, A.F. and Myerson, J.: *Discounting of Delayed Rewards: A Life-Span Comparison*. Psychological Science **5**(1), 33-36, 1994, http://dx.doi.org/10.1111/j.1467-9280.1994.tb00610.x,

[11] Sunstein, C.R.: *Default Rules Are Better Than Active Choosing (Often).* Trends in Cognitive Sciences **21**(8), 600-606, 2017, http://dx.doi.org/10.1016/j.tics.2017.05.003,

[12] Anderson, C.J.: *The psychology of doing nothing: Forms of decision avoidance result from reason and emotion*.
Psychological Bulletin **129**(1), 139-167, 2003,
http://dx.doi.org/10.1037/0033-2909.129.1.139,

[13] Hsee, C.K. and Hastie, R.: *Decision and experience: why don't we choose what makes us happy?*
Trends in Cognitive Sciences **10**(1), 31-37, 2006,
http://dx.doi.org/10.1016/j.tics.2005.11.007,

[14] Johansson, P.; Hall, L.; Sikström, S. and Olsson, *A.: Failure to detect mismatches between intention and outcome in a simple decision task*.
Science **310**(5745), 116-119, 2005,
http://dx.doi.org/10.1126/science.1111709,

[15] Soll, J.B.; Milkman, K.L. and Payne, J.W.: *A User's Guide to Debiasing*.
In: Keren, G.; Wu, G., eds.: *Wiley-Blackwell Handbook of Judgment and Decision Making*. Wiley Blackwell, Chichester, pp.924-952, 2015,

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

[17] Tolan, S.: *Fair and Unbiased Algorithmic Decision Making: Current State and Future Challenges*.
European Commission, Seville, 2018,

[18] Castelluccia, C. and Le Métayer, D.: *Understanding algorithmic decision-making: Opportunities and Challeges*.
European Parliamentary Research Service, Scientific Foresight Unit (STOA) PE 624.261, 2020,

[19] Alexander, V.; Blinder, C. and Zak, P.J.: *Why trust an algorithm? Performance, cognition, and neurophysiology*.
Computers in Human Behaviour **89**, 279-288, 2018,
http://dx.doi.org/10.1016/j.chb.2018.07.026,

[20] Lee, M.K.: *Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management*.
Big Data & Society **5**(1), 1-16, 2018,
http://dx.doi.org/10.1177/2053951718756684,

[21] Parasuraman, R. and Manzey, D.H.: *Complacency and Bias in Human Use of Automation: An Attentional Integration*.
Human Factors **52**(3), 381-410, 2010,
http://dx.doi.org/10.1177/0018720810376055,

[22] Lohr, S.: *Big Data Underwriting for Payday Loans*.
https://bits.blogs.nytimes.com/2015/01/19/big-data-underwriting-for-payday-loans,

[23] De Mayer, J.: *The use of big data and artificial intelligence in insurance*.
BEUC. The European Consumer Organisation, 2020,

[24] De Laat, P.B.: *Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability?*
Philosophy and Technology **31**(4), 525-541, 2018,
http://dx.doi.org/10.1080/03085147.2017.1412642,

[25] Kear, M.: *Playing the credit score game: algorithms, 'positive' data and the personification of financial objects*.
Economy and Society **46**(3-4), 346-368, 2017,
http://dx.doi.org/10.1080/03085147.2017.1412642,

[26]Lauer, J.: *Creditworthy: a history of consumer surveillance and financial identity in America*.
Columbia University Press, New York, 2017,

[27]Rosenblatt, E.: *Credit Data and Scoring. The First Triumph of Big Data and Big Algorithms.*
Academic Press, London, 2020,

[28] Christin, A.: *Algorithms in practice: Comparing web journalism and criminal justice*. Big Data & Society **4**(2), 1-14, 2017, http://dx.doi.org/10.1177/2053951717718855,

[29] Chiao, V.: *Fairness, accountability and transparency: notes on algorithmic decision-making in criminal justice*. International Journal of Law in Context **15**(2), 126-139, 2019, http://dx.doi.org/10.1017/S1744552319000077,

[30] Zweigl, K.A.; Wenzelburger, G. and Krafft, T.D.: *On Chances and Risks of Security. Related Algorithmic Decision Making Systems*. European Journal of Security Research **3**, 181-203, 2018, http://dx.doi.org/10.1007/s41125-018-0031-2,

[31]Harcourt, B.E.: *Against prediction: profiling, policing, and punishing in an actuarial age.* University of Chicago Press, Chicago, 2007,

[32] Kubler, K.: *State of urgency: Surveillance, power, and algorithms in France's state of emergency*. Big Data & Society **4**(2), 1-10, 2017, http://dx.doi.org/10.1177/2053951717736338,

[33] Bennett Moses, L. and Chan, J.: *Algorithmic prediction in policing: assumptions, evaluation, and accountability.* Policing and Society **28**(7), 806-822, 2018, http://dx.doi.org/10.1080/10439463.2016.1253695,

[34] Reich, A.: *Disciplined doctors: The electronic medical record and physicians' changing relationship to medical knowledge*. Social Science & Medicine **74**(7), 1021-1028, 2012, http://dx.doi.org/10.1016/j.socscimed.2011.12.032,

[35]Lyell, D. and Coiera, E.: *Automation bias and verification complexity: a systematic review.* Journal of the American Medical Informatics Association **19**(1), 121-127, 2016, http://dx.doi.org/10.1136/amiajnl-2011-000089,

[36] Bucher, T.: *The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms.* Information, Communication & Society **20**(1), 30-44, 2017, http://dx.doi.org/10.1080/1369118X.2016.1154086,

[37] Eslami, M., et al.: *'I always assumed that I wasn't really that close to [her]': Reasoning about Invisible Algorithms in News Feeds*. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. Association for Computing Machinery, Seoul, pp.153-162, 2015, http://dx.doi.org/10.1145/2702123.2702556,

[38]Eubanks, V.: *Automating Inequality. How high-tech tools profile, police, and punish the poor.* St. Martin's Press, New York, 2018,

[39] Barocas, S. and Selbst, A.D.: *Big Data's Disparate Impact*. California Law Review **104**(3), 671-732, 2016,

[40] De Visser, E.J., et al.: *A Little Anthropomorphism Goes a Long Way: Effects of Oxytocin on Trust, Compliance, and Team Performance With Automated Agents*. Human Factors **59**(1), 116-133, 2017, http://dx.doi.org/10.1177/0018720816687205,

[41] Hoff, K.A. and Bashir, M.: *Trust in automation integrating empirical evidence on factors that influence trust.* Human Factors **57**(3), 407-434, 2015, http://dx.doi.org/10.1177/0018720814547570,

[42] Burton, J.W.; Stein, M.K. and Jensen, T.B.: *A systematic review of algorithm aversion in augmented decision making*. Journal of Behavioural Decision Making **33**(2), 220-239, 2020, http://dx.doi.org/10.1002/bdm.2155,

[43] Ashton, A.H.; Ashton, R.H. and Davis, M.N.: *White-Collar Robotics: Levering managerial decision making*.
Managment Review **37**(1), 83-109, 1994,
http://dx.doi.org/10.2307/41165779,

[44] Jones, N.: *How machine learning could help to improve climate forecasts*.
Nature **548**(7668), 379-380 2017,
http://dx.doi.org/10.1038/548379a,

[45] Grove, W.M., et al.: *Clinical versus mechanical prediction: A meta-analysis*.
Psychological Assessment **12**(1), 19-30, 2000,
http://dx.doi.org/10.1037/1040-3590.12.1.19,

[46] Lee, J.D. and See, K.A.: *Trust in Automation: Designing for Appropriate Reliance*.
Human Factors **46**(1), 50-80, 2004,
http://dx.doi.org/10.1518/hfes.46.1.50_30392,

[47] Gillespie, T.: *The relevance of algorithms*.
In: Boczkowski, P.J.; Foot, K.A. and Gillespie, T., eds.: *Media Technologies: Essays on Communication, Materiality, and Society*. The MIT Press, Cambridge, pp.167-194, 2014,

[48] Christin, A.: *From Daguerreotypes to Algorithms: Machines, Expertise, and Three Forms of Objectivity*.
ACM SIGCAS Computers and Society **46**(1), 27-32, 2016,
http://dx.doi.org/10.1145/2908216.2908220,

[49] Akter, S., et al.: *Analytics-based decision-making for service systems: A qualitative study and agenda for future research*.
International Journal of Information Management **48**, 85-95, 2019,
http://dx.doi.org/10.1016/j.ijinfomgt.2019.01.020,

[50] Rader, E. and Wash, R.: *Trustworthy Algorithmic Decision-Making*.
http://bitlab.cas.msu.edu/trustworthy-algorithms,

[51] Castelfranchi, C. and Falcone, R.: *Trust theory: a socio-cognitive and computational model.*
John Wiley & Sons, Chichester, 2010,

[52] Tjøstheim, T.A.; Johansson, B. and Balkenius, C.: *A Computational Model of Trust-, Pupil-, and Motivation Dynamics.*
Proceedings of the 7th International Conference on Human-Agent Interaction. Association for Computing Machinery, Kyoto, pp.179-185, 2019,

[53] Alarcon, G.M.; Lyons, J.B. and Christensen, J.C.: *The effect of propensity to trust and familiarity on perceptions of trustworthiness over time*.
Personality and Individual Differences **94**, 309-315, 2016,
http://dx.doi.org/10.1016/j.paid.2016.01.031,

[54]DeSteno, D., et al.: *Detecting the Trustworthiness of Novel Partners in Economic Exchange.*
Psychological Science **23**(12), 1549-1556, 2012,
http://dx.doi.org/10.1177/0956797612448793,

[55] Fehr, E.; Fischbacher, U. and Kosfeld, M.: *Neuroeconomic Foundations of Trust and Social Preferences: Initial Evidence*.
American Economic Review **95**(2), 346-351, 2005,
http://dx.doi.org/10.1257/000282805774669736,

[56] Resnik, D.B.: *Scientific Research and the Public Trust.*
Science and Engineering Ethics **17**(3), 399-409, 2011,
http://dx.doi.org/10.1007/s11948-010-9210-x,

[57] Krueger, F. and Meyer-Lindenberg, A.: *Toward a Model of Interpersonal Trust Drawn from Neuroscience, Psychology, and Economics*.
Trends in Neurosciences **42**(2), 92-101, 2019,
http://dx.doi.org/10.1016/j.tins.2018.10.004,

[58] Hoffman, R.R.; Johnson, M.; Bradshaw, J.M. and Underbrink, A.: *Trust in Automation*.
IEEE Intelligent Systems **28**(1), 84-88, 2013,
http://dx.doi.org/10.1109/MIS.2013.24,

[59] De Visser, E.J., et al.: *Almost human: Anthropomorphism increases trust resilience in cognitive agents.*
Journal of Experimental Psychology: Applied **22**(3), 331-349, 2016,
http://dx.doi.org/10.1037/xap0000092,

[60] Efendić, E.; Van de Calseyde, P.P.F.M. and Evans, A.M.: *Slow response times undermine trust in algorithmic (but not human) predictions.*
Organizational Behavior and Human Decision Processes **157**, 103-114, 2020,
http://dx.doi.org/10.1016/j.obhdp.2020.01.008,

[61] Cummings, M.L.: *Automation bias in intelligent time critical decision support systems.*
AIAA 1st Intelligent Systems Technical Conference, American Institute of Aeronautics and Astronautics, Chicago, 2004,
http://dx.doi.org/10.2514/6.2004-6313,

[62] Bahner, J.E.; Hüper, A.-D. and Manzey, D.: *Misuse of automated decision aids: Complacency, automation bias and the impact of training experience.*
International Journal of Human-Computer Studies **66**(9), 688-699, 2008,
http://dx.doi.org/10.1016/j.ijhcs.2008.06.001,

[63] Zerilli, J.; Knott, A.; Maclaurin, J. and Gavaghan, C.: *Algorithmic Decision-Making and the Control Problem.*
Minds & Machines **29**(4), 555-578, 2019,
http://dx.doi.org/10.1007/s11023-019-09513-7,

[64] Villani, C.: *For a Meaningful Artificial Intelligence: Towards a French and European Strategy*.
Mission assigned by the Prime Minister Édouard Philippe, 2018,

[65] Wickens, C.D.; Clegg, B.A.; Vieane, A.Z. and Sebok, A.L.: *Complacency and Automation Bias in the Use of Imperfect Automation.*
Human Factors **57**(5), 728-739, 2015,
http://dx.doi.org/10.1177/0018720815581940,

[66] Wesley, D. and Dau, L.: *Complacency and Automation Bias in the Enbridge Pipeline Disaster.*
Ergonomics in Design **25**(1), 17-22, 2017,
http://dx.doi.org/10.1177/1064804616652269,

[67] Skitka, L.J.; Mosier, K. and Burdick, M.D.: *Accountability and automation bias.*
International Journal of Human-Computer Studies **52**(4), 701-717, 2000,
http://dx.doi.org/10.1006/ijhc.1999.0349,

[68] Önkal, D., et al.: *The Relative Influence of Advice From Human Experts and Statistical Methods on Forecast Adjustments*.
Journal of Behavioural Decision Making **22**(4), 390-409, 2009,
http://dx.doi.org/10.1002/bdm.637,

[69] Dietvorst, B.J.; Simmons, J.P. and Massey, C.: *Algorithm aversion: People erroneously avoid algorithms after seeing them err.*
Journal of Experimental Psychology: General **144**(1), 114-126, 2015,
http://dx.doi.org/10.1037/xge0000033,

[70] Prahl, A. and Van Swol, L.: *Understanding algorithm aversion: When is advice from automation discounted?*
Journal of Forecasting **36**(6), 691-702, 2017,
http://dx.doi.org/10.1002/for.2464,

[71] Logg, J.M.; Minson, J.A. and Moore, D.A.: *Algorithm Appreciation: People Prefer Algorithmic to Human Judgment.*
Organizational Behavior and Human Decision Processes **151**, 90-103, 2019,
http://dx.doi.org/10.1016/j.obhdp.2018.12.005,

[72] Yeomans, M.; Shah, A.; Mullainathan, S. and Kleinberg, J.: *Making sense of recommendations.*
Journal of Behavioral Decision Making **32**(4), 403-414, 2019,
http://dx.doi.org/10.1002/bdm.2118,

[73] Arya, S.; Eckel, C. and Wichman, C.: *Anatomy of the credit score*. Journal of Economic Behaviour & Organization **95**, 175-185, 2013, http://dx.doi.org/10.1016/j.jebo.2011.05.005,

[74] Edelman, D.: *Credit this: how the banks decide your credit score*. Significance **5**(2), 59-61, 2008, http://dx.doi.org/10.1111/j.1740-9713.2008.00287.x,

[75] Schäufele, F.: *Profiling zwischen sozialer Praxis und technischer Prägung.* Springer, Wiesbaden, 2017,

[76] Poon, M.: *Scorecards as Devices for Consumer Credit: The Case of Fair, Isaac & Company Incorporated.* The Sociological Review **55**(2), 284-306, 2007, http://dx.doi.org/10.1111/j.1467954X.2007.00740.x,

[77] Aitken, R.: *'All data is credit data': Constituting the unbanked.* Competition & Change **21**(4), 274-300, 2017, http://dx.doi.org/10.1177/1024529417712830,

[78] Hurley, M. and Adebayo, J.: *Credit Scoring in the Era of Big Data*. Yale Journal of Law and Technology **18**(1),148-216, 2017,

[79] Wei, Y.; Yildirim, P.; Van den Bulte, C. and Dellarocas, C.: *Credit Scoring with Social Network Data.* Marketing Science **35**(2), 234-258, 2016, http://dx.doi.org/10.1287/mksc.2015.0949,

[80] Saif, M.A.; Prisyazhny, A.V. and Medvedeva, M.A.: *On the Model of Credit Score Calculation Using Social Networks Data.* Marketing Science **35**(2), 234-258, 2018, http://dx.doi.org/10.1063/1.5044043,

[81] Deville, J. and van der Velden, L.: *Seeing the invisible algorithm: the practical politics of tracking the credit trackers.* In: Amoore, L. and Piotukh, V., eds.: *Algorithmic life. Calculative devices in the age of big data*. Routledge, New York, pp.87-106, 2016, http://dx.doi.org/10.4324/9781315723242,

[82] Lohokare, J.; Dani, R. and Sontakke, S.: *Automated data collection for credit score calculation based on financial transactions and social media*. International Conference on Emerging Trends & Innovation in ICT. IEEE, Pune, pp.134-138, 2017,

[83] Siddiqi, N.: *Intelligent Credit Scoring. Building and Implementing Better Credit Risk Scorecards*. Wiley, Hoboken, 2017,

[84]Springer, A.: *Accurate, Fair and Explainable: Building Human-Centred AI*. Ph.D. Thesis. UC Santa Cruz, Santa Cruz, 2019,

[85] Cohen, P.: *In Lending Circles, a Roundabout Way to a Higher Credit Score*. New York Times, 2014, http://www.nytimes.com/2014/10/11/your-money/raising-a-credit-score-from-zero-to-789-in-26-months.html,

[86] Strle, T.: *Looping minds: How cognitive science exerts influence on its findings.* Interdisciplinary Description of Complex Systems **16**(4), 533-544, 2018, http://dx.doi.org/10.7906/indecs.16.4.2,

[87] Strle, T. and Markič, O.: *Looping effects of neurolaw and the precarious marriage between neuroscience and the law*. Balkan Journal of Philosophy **10**(1), 17-26, 2018, http://dx.doi.org/10.5840/bjp20181013,

[88] Strle, T.: *The Image of Bounded Rationality and Feedback Effects of Modifying Choice Environments*. Cognitive Science: Proceedings of the 22nd International Multiconference Information Society – IS 2019. Institut Jožef Stefan, Ljubljana, pp.56-60, 2019,

[89] Smith, A.: *Public Attitudes Toward Computer Algorithms*.
http://www.pewresearch.org/internet/2018/11/16/attitudes-toward-algorithmic-decision-making,
[90] Taylor, A. and Sadowski, S.: *How Companies Turn Your Facebook Activity Into a Credit Score.*
http://www.thenation.com/article/archive/how-companies-turn-your-facebook-activity-credit-score,
[91] Xu, W.: *Toward Human-Centred AI: A Perspective from Human-Computer Interaction*.
Interactions **26**(4), 42-46, 2019,
http://dx.doi.org/10.1145/3328485,
[92] Lee, M.K. and Baykal, S.: *Algorithmic Mediation in Group Decisions: Fairness Perceptions of Algorithmically Mediated vs. Discussion-Based Social Division*.
Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing – CSCW. Association for Computing Machinery, Portland, 2017,
[93] Dreyer, S. and Schulz, W.: *The General Data Protection Regulation and Automated Decision-making: Will it deliver? Potentials and limitations in ensuring the rights and freedoms of individuals, groups and society as a whole.*
Bertelsmann Stiftung, 2019,
http://dx.doi.org/10.11586/2018018,
[94] Castets-Renard, C.: *Accountability of Algorithms in the GDPR and Beyond: A European Legal Framework on Automated Decision-Making*.
Fordham Intellectual Property, Media and Entertainment Law Journal **30**(1), 91-137, 2019.

Interdisciplinary Description of Complex Systems 19(4), 561-579, 2021

A NOVEL DISCRETE INTERNAL MODEL CONTROL METHOD FOR UNDERACTUATED SYSTEM

Imen Saidi[1,] *, Islem Bejaoui[1] and Maria Gabriella Xibilia[2]

[1]University of Tunis El Manar, U.R.LARA Automatique, Ecole Nationale d'Ingénieurs de Tunis
Tunis, Tunisia

[2]Università degli studi di Messina, Department of Engineering
Messina, Italy

DOI: 10.7906/indecs.19.4.8
Regular article

Received: 24 January 2021.
Accepted: 10 May 2021.

ABSTRACT

This article provides a comparative analysis of two common control configurations used to control the side-stream distillation used to separate benzene, toluene and xylene as suggested by Doukas and Lyben. Their under-actuated model has been considered as the model of distillation column and the internal model controller is designed considering a Singular Value Decomposition (SVD) and Virtual Inputs (VI) techniques. An internal controller design based on VI is proposed in this article for this kind of underactuated systems. This design is used to control in parallel the distillation process and its model in real time. The proposed controller design is simple and systematic in relation with the desired closed loop specifications of the internal model control structure. Furthermore, the controller obtained ensure robustness to process variations. The SVD technique can realize the decoupling of under-actuated processes and wipe out unrealizable factors by introducing compensation terms, affecting the dynamic of the system. The aim of this article is to make a comparison between our proposed VI controller and the SVD approach. The results we obtained confirmed the potentials of the proposed controller based on VI considering the set point tracking and its robustness.

KEY WORDS

under-actuated systems, virtual input technique, singular value decomposition technique, internal model control, stability

CLASSIFICATION

ACM: D.1.0, I.6.0, J.2
JEL: L64
PACS: 02.30.Yy, 89.20.Kk

*Corresponding author, η: imen.saidi@gmail.com; +216 97 73 4956;
Campus Universitaire Farhat Hached el manar bp 37, le Belvedere 1002 Tunis

INTRODUCTION

The contribution of this article is study of two different control strategies for a class of chemical process industries. The model of the distillation column used in this plant, which is characterized by an underactuated structure. According to the diversity and complexity of these systems, it is important to emphasize that none of the technique proposed and developed for fully actuated systems can be applied directly to any underactuated system. Therefore, it is meaningful to develop control methods for this class of systems, and more precisely to develop the most optimized controller design.

The control objective for systems characterized by the fact that there are more degrees of freedom than actuators, is to obtain a desirable behavior of several output variables by simultaneously manipulating several inputs channels. Under-actuated systems are less sensitive to modelling errors, so it has to be controlled in its original form to obtain robust stability and performance [1].

Garcia and Morari [2], presented the concept of Internal Model Control (IMC) and have already proved the effectiveness of the framework for robust control of different kinds of Singular-Input Singular Output (SISO) systems. The IMC structure is composed of three principal parts: the process, the internal model in parallel to the process and the controller, which is the dynamic inversion of the model. Recently, many methods have been proposed to control MIMO systems with time delays. For example, Zhang and Pang [3] proposed a Closed-loop Gain Shaping Algorithm (CGSA) using Padé approximation, which is sufficiently accurate in view of stability analysis. Jin et al [4] introduced a design method of decoupling internal model control; the basic idea is to realize the decoupling of the controller of non-square processes by inserting some compensation Relative Normalized Gain Array (RNGA). An equivalent transfer function matrix is introduced to approximate the pseudo-inverse of the process transfer function matrix, which makes the design of decoupling internal model control simple and easy to calculate.

Shan and Wang [5] integrated IMC with disturbance controllers by choosing different forms of external input/output disturbance. When this disturbance is applied directly as input to the controller, the design of controller needs to compensate the effect of slow dynamic poles by adding some constraints. Pamela et al. [6] introduced an approach to regulate the heater power in those systems, which must control the temperature in food processing, pharmaceuticals and in polymerization. The objective is to control the system with both PI controller and IMC structure and to analyze its performance parameters. Jin et al. [7] proposed a novel design IMC controller based on Singular Value Decomposition (SVD), this approach uses SVD in the inverse of the steady-state gain matrix of process. The last decades have shown an increasing interest in the control of under-actuated systems, many IMC methods can be introduced to achieve considerable results on the control for these kinds of systems. For this class of system, the number of inputs is smaller than the number of outputs, which means that the transfer function matrix is not square, and the issue of inversion exists. As a result, the problem of inversion was solved in many literature researches previously mentioned [4-6] by decoupling the internal model control based on the Relative Normalized Gain Array (RNGA), Singular Value Decomposition (SVD) and an Equivalent Transfer Function (ETF) matrix. The solution brought by these methods require complicated calculations and many instructions to implement [7].

Otherwise, our proposed approach; Virtual Inputs is used to augment the system inputs inserting a certain number of virtual columns to the model in order to make it a square matrix. These virtual columns have no influence on the response of the system, and they will be eliminated

after. The control of under-actuated systems is challenging, hence the necessity to identify the control technique with less interactions and better performances such as overshoot, setting time etc. Based on internal model control for discrete under-actuated systems a comparison based on the inversion technique of the model will be made between Singular Value Decomposition (SVD) and Virtual Inputs (VI) to realize the internal model controller.

In this article, section 2 summarizes the fundamental control problem to be solved. Section 3 describe the general IMC structure. Section 4 details the design of the internal controller based on Singular Value Decomposition. Section 5 details the design of the internal controller based on Virtual Inputs. At the end, we show how our proposed methodology based on Virtual Inputs at distillation column could give good results compared to the other method described previously.

PROBLEM FORMULATION

The process to be controlled is assumed linear and discrete-time governed by the following equation [8]

$$\begin{cases} x(k+1) = Fx(k) + \sum_{i=1}^{n}\sum_{j=1}^{m} Hu(k-\tau_{ij}) \\ y(k) = Cx(k) \end{cases} \tag{1}$$

Wheren $x \in \mathbb{R}^n$ i is the system state, $u \in \mathbb{R}^m$ is the vector of manipulated variables, $y \in \mathbb{R}^n$ is the vector of system outputs, t_{ij} is the time delays, F, H and C are matrices of corresponding dimensions [8]. For the IMC design technique, it is convenient to express this model in frequency response form. Taking Z transforms of (1), the following input-output model is obtained:

$$y(z) = C(zI - F)^{-1}\sum_{i=1}^{n}\sum_{i=1}^{n} Hz^{-\tau_{ij}}\, u(z) = G(z)\, u(z) \tag{2}$$

Where $G(z)$ is the system transfer matrix of dimension $(n \times m)$, the number of control inputs is equal to m, the number of outputs is equal to n, making it a rectangular matrix, which has the form

$$G(z) = z^{-1}\begin{pmatrix} z^{-\tau_{11}} g_{11}(z) & z^{-\tau_{12}} g_{12}(z) & \dots & z^{-\tau_{1m}} g_{1m}(z) \\ z^{-\tau_{21}} g_{21}(z) & z^{-\tau_{22}} g_{22}(z) & \dots & z^{-\tau_{2m}} g_{2m}(z) \\ \vdots & \vdots & \ddots & \vdots \\ z^{-\tau_{n1}} g_{n1}(z) & z^{-\tau_{n2}} g_{n2}(z) & \dots & z^{-\tau_{nm}} g_{nm}(z) \end{pmatrix} \tag{3}$$

Where the elements $g_{ij}(z)$, are the transfer functions of z and t_{ij} is the delay in the response of output i to input j. The synthesis of an IMC controller that is equal to the inverse of the model expression and is fundamental to ensure perfect set-point tracking and this represents the basic problem of the IMC approach. In fact, the realization of the direct model inverse is difficult, and sometimes not possible to realize, for many physical systems [8, 9]. This perfect controller cannot be implemented for the following reasons.

1.) If the model contains time delays, its inverse involves predictive terms, which make the controller unrealizable.

2.) If the zeros of the transfer matrix of the model outside the complex unit circle yield an unstable perfect controller.
3.) System equipped with a perfect controller is extremely sensitive to modelling errors and time delays.
4.) The direct model inversion is also impossible in the case of underactuated systems.

In fact, the model must provide an accurate description of the process dynamics and characteristics. Therefore, the model expression must be very close to that of the plant. For underactuated systems, the number of control inputs is less than the number of outputs and therefore we will have a rectangular matrix that is invertible. This represents the major problem encountered. In this article, to resolve this problem, it is proposed to develop some methods of inversion in the case of under-actuated systems.

INTERNAL MODEL CONTROL DESCRIPTION

The development of the IMC structure has progressed in recent years in order to design an optimal feedback controller. In this section we present the general IMC structure and we describe its basic principles and properties. Due to the fact that the traditional IMC methods cannot solve the control problem of non-square systems, we introduce two design technique to realize the internal controller but, when using a matrix to describe a non-square system, the issue of inversion often emerges.

THE GENERAL IMC DESCRIPTION

The general IMC structure of multivariable systems adopted in this article is shown in Figure 1, where $y(z)$, $y_m(z)$ and $u(z)$ are the output vector of the process, the output of internal model and the control variable, respectively. *r(z)* is the set-point vector, *d(z)* is the disturbance, $G(z)$ and $M(z)$ represent the transfer function matrix of process and its model, $C(z)$ is the transfer matrix of the IMC controller.

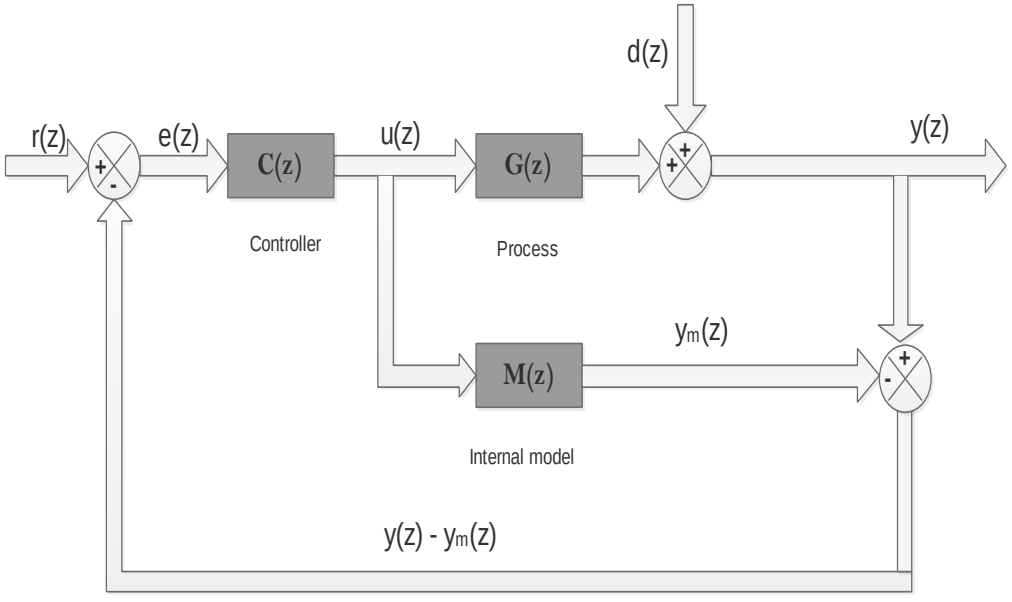

Figure 1. The general IMC structure.

We begin by reviewing the properties of the IMC structure. This structure is equivalent to a conventional feedback loop with controller. From Figure 1 the inputs vector *u(z)* and the system outputs vector *y(z)* are expressed by

$$u(z) = [I + C(z)(G(z) - M(z))]^{1} C(z)(r(z) - d(z)) \tag{4}$$

$$y(z) = d(z) + C(z)G(z)[I + C(z)(G(z) - M(z))]^{1} (r(z) - d(z)) \tag{5}$$

Property 1. Dual Stability; Assuming an ideal model $M(z) = G(z)$ and $d(z) = 0$, stability of both the controller C(z) and the process $G(z)$ is then sufficient for overall system. Then equations (4) and (5) becomes:

$$u(z) = C(z)r(z) \tag{6}$$

$$y(z) = C(z)G(z)r(z) \tag{7}$$

Therefore, the system poles as well as the controller poles must lie inside the unit circle (UC) for stability. On the other hand, when $G(z) = M(z)$ the stability is not affected by adding constraints on the inputs.

Property 2. Perfect Control; Assume that the controller $C(z) = M(z)^{-1}$, yields a closed-loop stable IMC loop, this controller is equivalent to the inverse model to achieves perfect set-point satisfaction despite any disturbance. Furthermore, $M(z)^{-1}$ is often not realizable. When $C(z) = M(z)^{-1}$ is verified, transfer function (7) becomes

$$y(z) = M^{-1}(z)G(z)r(z) = r(z) \tag{8}$$

Property 3. Zero Offset; Assume that the steady state gain controller is equal to the inverse model gain $C(1) = M(1)^{-1}$ and the closed-loop system in Figure 1 is stable.

The key to apply the IMC structure is the controller who would yield the best output response possible. However, as in the under-actuated case, this perfect controller cannot be implemented for previous reasons. We will discuss in the next section the controller design using two different methods; Singular Value Decomposition (SVD) and Virtual Input (VI). These methods are based on the transfer function matrix of the model of the process.

IMC STRUCTURE BASED ON SINGULAR VALUE DECOMPOSITION

The application of the internal model structure to under-actuated systems is considered like our main target. In this section, we will describe the design phase of the internal controller based on SVD and the implementation at the level of the structure IMC. This approach of design can realize the decoupling of under-actuated processes and eliminate the unrealizable factors by inserting compensated terms. Meanwhile, a non-diagonal filter is designed based on SVD matrix theory. We discuss with more details this approach in the following.

The IMC structure of an under-actuated systems based on SVD is shown in Figure 2, where $C_{SVD}(z)$ is the IMC controller.

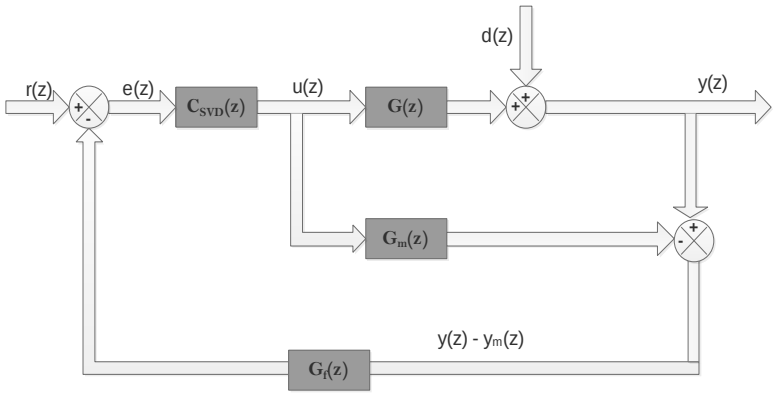

Figure 2. IMC structure based on SVD method.

The SVD of the transfer matrix *G(z)* of the under-actuated systems is as follows:

$$G(z) = U.\Sigma.V^{H} \tag{9}$$

Where, U is an $(m' \ m)$ orthonormal matrix and $U'U = I_m$, (I_m is the identity matrix), $\sum$ is an diagonal matrix with singular values of $G(z)$ in the diagonal, V^H is an $(n \times m)$ orthonormal matrix and $V'V = I_m$ and H represented the conjugate transpose. The SVD is essentially trying to reduce the rank of the matrix and to approximate them as a linear combination [10]. The closed loop transfer function matrix deduced from Figure 2 is given by:

$$H(z) = G(z)C_{SVD}(z) \times \left[I_m + (G(z) - G_m(z)) \times C_{SVD}(z)\right]^{-1} \tag{10}$$

where $C_{SVD}(z)$ is a non-square internal model controller based on Singular Value Decomposition has two main roles; on one hand make the internal control, while compensating and decoupling the system to reduce coupling between channels and, on the other hand, it can satisfy the robustness and the performances of the systems. The design method of the under-actuated internal model controller $C_{SVD}(z)$ is given below.

In the nominal case, $G(z) = G_m(z)$ equation (10) becomes

$$H(z) = G(z)C_{SVD}(z) = G_m(z)C_{SVD}(z) = \text{diag}\{h_{ii}(z)\} \tag{11}$$

Where, $h_{ii} \neq 0$, $i = 1, 2, \ldots\ldots, m$. Hence, in the traditional IMC structure [11], the closed loop transfer function $H(z)$ and the general decoupled internal model controller $C_{IMC}(z)$ should be respectively

$$H(z) = G_m(z)C_{IMC}(z) \tag{12}$$

$$C_{IMC}(z) = G_m^{-1}(z)H(z) \tag{13}$$

It is obvious that once $G_m^{-1}(z)$ and $H(z)$ are specified, it is possible to determine $C_{IMC}(z)$. We start by finding the $G_m^{-1}(z)$ but in under-actuated systems; the exact inverse of the model does not exist, so we replace it with the generalized inverse of $G_m(z)$ [12]. Hence, the result of this inversion is the unit matrix $G_m^{*}(z)$:

$$\mathrm{G}_m^*(\mathrm{z}) = \mathrm{G}_m^H(\mathrm{z})\left[G_m G_m^H\right]^{-1} \tag{14}$$

Where $G_m^*(\mathrm{z})$ and $G_m^H(z)$ are the pseudo-inverse and the Hermitian matrix of $G_m(\mathrm{z})$ respectively.

The expression (13) becomes

$$C_{IMC}(\mathrm{z}) = G_m^*(\mathrm{z})\,\mathrm{H}(\mathrm{z}) = \mathrm{G}_m^H(\mathrm{z})\left[G_m G_m^H\right]^{-1} H(\mathrm{z}) \tag{15}$$

Next, we choose the appropriate $H(z)$. Whereas the model $G_m(\mathrm{z})$ consists into two parts to handle many limitations in this technique

$$G_m(\mathrm{z}) = G_{m-}(\mathrm{z})G_{m+}(\mathrm{z}) \tag{16}$$

Where $G_{m+}(z)$ contains time delays and zeros of $G_m(z)$ outside the unit circle such $G_{m-}(z)$ has a stable and realizable inverse.

In the traditional internal model controller design [12], adding the filter is used to supplement the model mismatch and ignore the error caused by the non-minimum phase portion, as shown in equation (17).

$$C_{IMC}(\mathrm{z}) = G_m^{-1}(\mathrm{z})F(\mathrm{z}) \tag{17}$$

Substituting equations (16) and (17) into equation (12) leads to:

$$H(\mathrm{z}) = G(\mathrm{z})\mathrm{C}_{IMC}(\mathrm{z}) = \mathrm{G}_m(\mathrm{z})\mathrm{C}_{IMC}(\mathrm{z}) = \mathrm{G}_{m+}(\mathrm{z})\mathrm{G}_{m-}(\mathrm{z})G_m^{-1}(\mathrm{z})F(\mathrm{z}) = \mathrm{G}_{m+}(\mathrm{z})F(\mathrm{z}) \tag{18}$$

Where $F(z)$ is a designed filter of internal model controller. It can be considered as the general form [13]:

$$F(\mathrm{z}) = \frac{b_f z^{-1}}{1 - a_f z^{-1}} \tag{19}$$

Among them:

$$a_f = e^{(\frac{T}{\mathrm{l}})} \tag{20}$$

$$b_f = 1 - a_f \tag{21}$$

In the above formula, λ is the time constant of the design filter, it determines the bandwidth of the closed-loop system and thus serves as a tuning parameters for performance and robustness, T is the sampling time of the system. Furthermore, when the high performance is required, $\mathrm{F(z)} \approx I_m$ but it is intuitively obvious that this choice makes the system very sensitive and it can very easily become unstable even for small modeling errors. In addition, the completely decoupling can be obtained by choosing $G_{m+}(z)$ and $F(z)$ diagonal. Even when dynamic interactions are allowed, $G_{m+}(\mathrm{z})$ and $F(z)$ must satisfy [13]:

$$G_{m+}(1) = I_m,\ F(1) = I_m \tag{22}$$

In the following discussion, we give procedures for finding $G_{m+}(z)$ and a rule for filter design. Assume $G(z) = G_m(z)$, and that $G_{m+}(z)$ is diagonal with the following form [14]

$$G_{m+}(z) = diag\left\{ e^{-\tau_{ij} Tz} \prod_{P=1}^{Q_i} \left\langle \frac{-z+z_p}{-z+\bar{z}_p} \right\rangle^{ri} , i = 1,2,.....,m \right\} \tag{23}$$

where t_{ij}^{T} is the maximum prediction in the i-th row of $G_m^*(z)$, z_p is the pole in outside of the unit circle, $\bar{z}_p$ is the conjugate complex of z_p, r_i represents the maximum number of the same pole of $G_m^*(z)$.

IMPROVEMENT OF FILTER

According to the equation (17), the robustness of the $C_{SVD}(z)$ often cannot meet the requirement, for solving this problem, this approach need to improve the filter *F(z)*. The filtering structure is designed in general, which makes control system bear the capacity of high-dimensional decoupling and fast response. First, we use the SVD in the inverse of the transfer functions matrix of process $G(z)$. Then, we can use the term after decomposition to improve the filter $F(z)$ and to obtain the controller based on SVD.

Step 1: Use the SDV in the inverse of the steady-state gain matrix of $G(z)$

$$\left[G_m(z=1)\right]^* = U\Sigma V^H \tag{24}$$

Step 2: Let W_v satisfy the following formula

$$V^H W_V = I \tag{25}$$

Step 3: The improved internal model controller is as follows

$$C_{SDV}(z) = G_m^*(z) G_{m+}(z) W_v F(z) W_v^{-1} \tag{26}$$

The robustness of the system can be greatly enhanced by adding a filter $G_f(z)$ in the feedback loop, and the filter time constant can be set to half of the maximum delay time in this loop [15].

IMC STRUCTURE BASED ON VIRTUAL INPUTS (VI)

Focusing now on the inversion method of the under-actuated systems based on Virtual Inputs. To successfully apply our approach, firstly we need to modify the IMC basic structure mentioned in Figure 1, so that it becomes applicable to underactuated systems with more outputs than control inputs. Secondly, we design an approximate inverse of the model plant which is inspired by the studies of [16, 17] in the case of MIMO systems and over-actuated systems [18].

The modified IMC structure presented in Figure 3 is characterized by two more blocks with respect to the basic IMC structure presented in Figure 1 [19]. The first new block is the Virtual Inputs Augmentation VIA (z) which is used to augment the system inputs inserting virtual ($n \times (n - m)$) column to the transfer matrix of the non-square system in order to make it of dimension ($n \times m$), so that it can be inverted. The second one is the Virtual Inputs Removing VIR (z) block which is used to eliminate the exceeding virtual inputs [19].

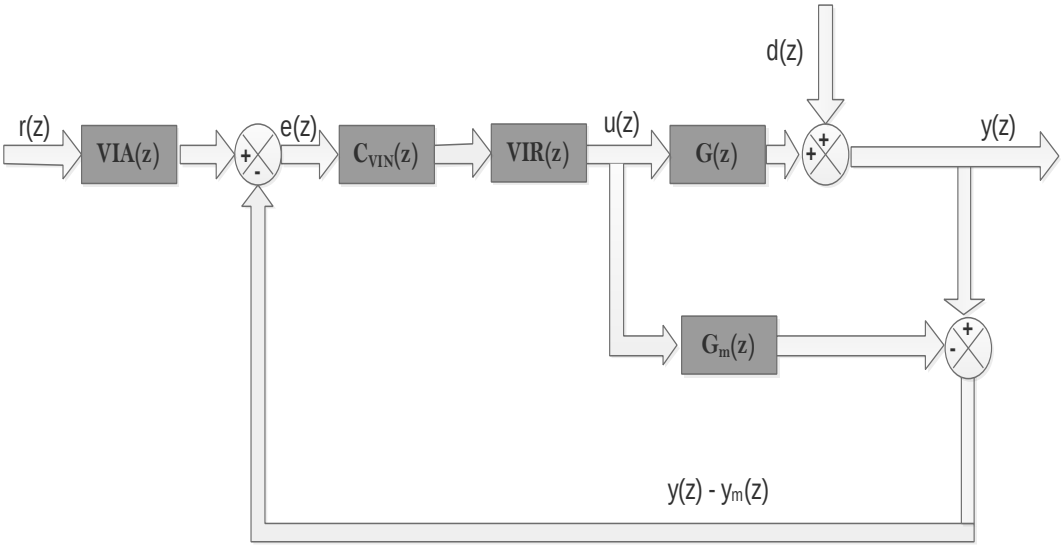

Figure 3. IMC structure based on Virtual Inputs method.

The inserting columns mentioned previously can be chosen as first-order transfer functions, which verify the stability criterion, and in order to simplify the study and avoid inversion problems [19].

The studied system can be shown through the following equation [18, 19]:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} G_{11}(z) & G_{12}(z) & \cdots & G_{1m}(z) \\ G_{21}(z) & G_{22}(z) & \ldots & G_{2m}(z) \\ \vdots & \vdots & \ddots & \vdots \\ G_{n1}(z) & G_{n2}(z) & \ldots & G_{nm}(z) \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{bmatrix} \tag{27}$$

When the model and the system match perfectly $G(z) = G_m(z)$, the augmented system $\breve{G}(z)$ and its exact model are expressed as:

$$\tilde{G}(z) = \tilde{G}_m(z) = \underbrace{\begin{bmatrix} G_{11} & \ldots & G_{1m} \\ \vdots & \ddots & \vdots \\ G_{n1} & \ldots & G_{nm} \end{bmatrix}}_{\substack{Initial\ transfer\ function \\ (n \times m)}} \underbrace{\begin{bmatrix} G_{1m+1} & \ldots & G_{1n} \\ \vdots & \ddots & \vdots \\ G_{nm+1} & \ldots & G_{nn} \end{bmatrix}}_{\substack{Added\ transfer\ function \\ (n \times (n-m)}} \tag{28}$$

The Virtual Inputs Controller $C_{VIN}(z)$ design presented in Figure 4 is based on the inversion method reported in [18, 19].

The closed-loop transfer function matrix $C_{VIN}(z)$ between $e(z)$ and $u(z)$ is derived as:

$$C_{VIN}(z) = K_2 (I_n + K_1 \tilde{G}_m(z))^{-1} K_1 \tag{29}$$

Where K_1 is an invertible square matrix $(n \times n)$ used to ensure the stability of the controller and it can be expressed by $K_1 = \alpha I_m, \alpha \in R^+$, a second gain K_2 is introduced as reported in Figure 4 in order to compensate the static errors. The gain K_2 is given by equation (30).

$$K_2 = K_1^{-1} (I_m + K_1 \tilde{G}_m(1))^{-1} \tilde{G}_m(1)^{-1} \tag{30}$$

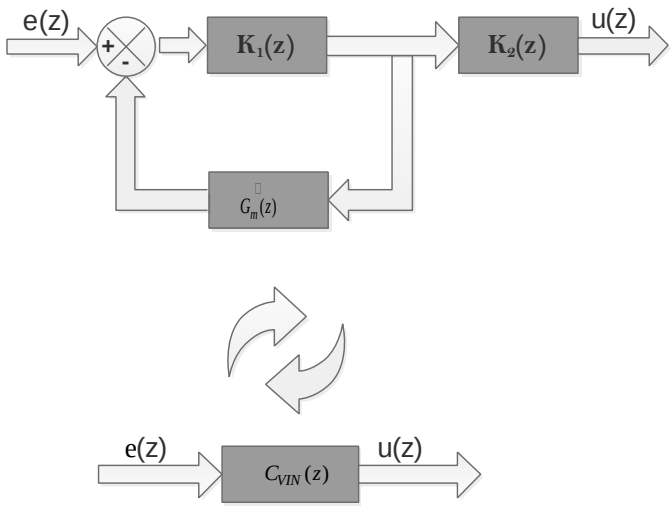

Figure 4. Structure of the Virtual Input controller.

For sufficiently high values of a , the controller $C_{VIN}(z)$ approaches the inverse of internal model $\check{G}_m(z)$:

$$C_{VIN}(z) \approx \tilde{G}_m(z)^{-1} \tag{31}$$

For the construction of the controller, it is necessary to respect the fact that:

$$C_{VIN}(z).\tilde{G}_m(z) = \mathrm{I}_n \tag{32}$$

The virtual inputs method proposed for the design of the controller by internal model is valid whatever the number of outputs $n \in \mathbb{N}^*$ and inputs $m \in \mathbb{N}^*$ of a physical system. In fact, this approach is applicable for the following classes of systems:

- Monovariable system where $n = m = 1$,
- Overactuated system where $n < m$,
- Underactuated system where $n > m$.

SIMULATION RESULTS AND COMPARISON

To analyze the comparative study on the above control technique, we considered a chemical process industry. In order to test the control effect of discrete underactuated internal model controller based on SVD and VI, a side-stream distillation control problem suggested by Doukas and Lyben will be used [20].

Figure 5 shows the side-stream distillation scheme that serves to separate benzene, toluene and xylene. The problem posed in this case was to control the concentrations of four impurities in three product streams with only three manipulated variables: reboiler duty, reflux ratio and side stream flow rate.

Doukas and Lyben [20] treated the distillation column as a 4 × 3system. An internal controller should be design for the under-actuated system. Doukas and Lyben model can be represented as

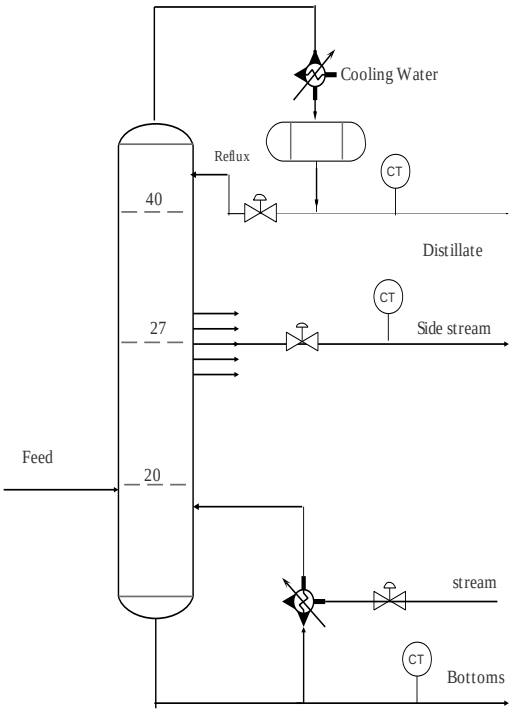

Figure 5. Schematic diagram of distillation column [20].

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} G_{11}(z) & G_{12}(z) & G_{13}(z) \\ G_{21}(z) & G_{22}(z) & G_{23}(z) \\ G_{31}(z) & G_{32}(z) & G_{33}(z) \\ G_{41}(z) & G_{42}(z) & G_{43}(z) \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \tag{33}$$

where the toluene in bottom *(y_1)*, toluene in bottom *(y_2)*, benzene in side draw *(y_3)* and benzene in side draw *(y_4)* are the four controlled variables. The reboiler duty *(u_1)*, reflux ratio *(u_2)* and side draw *(u_3)* are the manipulated variables [20].

The $(4' 3)$ transfer function matrix presented in equation (33) as follows

$$G_{11}(z) = \frac{-5.12\,z - 0.61}{z - 0.41} z^{-1}, G_{12}(z) = \frac{0.0017\,z^2 + 0.0356\,z + 0.0117}{z^2 - 1.27\,z + 0.40} z^{-1} \tag{34}$$

$$G_{13}(z) = \frac{-0.3811\,z^2 - 1.078\,z - 0.07728}{z^2 - 1.263\,z + 0.3985} z^{-1}, G_{21}(z) = \frac{2.507z + 0.5044}{z - 0.4967} z^{-1} \tag{35}$$

$$G_{22} = \frac{-0.0175\,z^2 - 0.0208\,z - 0.0566}{z^2 - 1.72\,z + 0.74} z^{-1}, G_{23} = \frac{0.0016\,z + 0.0015}{z^2 - 1.95\,z + 0.95} z^{-6} \tag{36}$$

$$G_{31}(z) = \frac{2.35\,z^2 + 0.0182\,z + 0.17}{z^2 - 0.0018\,z + 0.0077} z^{-1}, G_{32}(z) = \frac{0.0077\,z^2 + 0.003837\,z + 0.0111}{z^2 - 0.49\,z + 0.0607} z^{-1} \tag{37}$$

$$G_{33} = \frac{-0.29\,z^2 - 0.0225\,z - 0.0091}{z^2 - 0.0299\,z + 0.0002} z^{-1}, G_{41} = \frac{-5.66\,z - 0.87}{z - 0.44} z^{-1} \tag{38}$$

$$G_{42} = \frac{-0.0706\,z^2 - 0.0324\,z - 0.0166}{z^2 - 0.46\,z + 0.0551} z^{-1}, G_{43} = \frac{0.94\,z^2 + 0.63\,z + 0.0008}{z^2 - 0.81\,z + 0.16} z^{-1} \tag{39}$$

In this section, we evaluate the controller performance using the above-mentioned approach.

SIMULATION RESULTS USING THE SVD CONTROLLER

Starting with the internal controller based on Singular Value Decomposition, and using the whole approach seen in section 4. Using equation (23), $G_{m+}(z)$ can be obtained by

$$G_{m+}(z) = diag\left\{\frac{1}{z-7.75}, \quad \frac{1}{z-60}, \quad \frac{1}{z-0.59}, \quad \frac{1}{z-1.91}\right\} \tag{40}$$

Using equation (24), L, S ,V is given as follows

$$U = \begin{bmatrix} -0.4523 & -0.4230 & 0.7841 & 0.0399 \\ -0.8883 & 0.1387 & -0.4377 & 0.0022 \\ 0.0327 & -0.1356 & -0.1043 & 0.9847 \\ -0.0720 & 0.8851 & 0.4274 & 0.1696 \end{bmatrix} \tag{41}$$

$$\sum = \begin{bmatrix} 0.6198 & 0 & 0 \\ 0 & 0.0902 & 0 \\ 0 & 0 & 0.0553 \end{bmatrix} \tag{42}$$

$$V = \begin{bmatrix} 0.0243 & -0.5117 & -0.8588 \\ 0.9968 & -0.0534 & 0.0600 \\ 0.0765 & 0.8575 & -0.5088 \end{bmatrix} \tag{43}$$

Using equation (25), W_v, W_v^{-1} are given as follows

$$W_v = \begin{bmatrix} 0.0243 & 0.9968 & 0.0765 \\ -0.5117 & -0.0534 & 0.8575 \\ -0.8588 & 0.0600 & -0.5088 \end{bmatrix} \tag{44}$$

$$W_v^{-1} = \begin{bmatrix} 0.0243 & -0.5117 & -0.8588 \\ 0.9967 & -0.533 & 0.0600 \\ 0.0766 & 0.8575 & -0.5087 \end{bmatrix} \tag{45}$$

Considering the realization of the controller, the added filter $F(z)$ is

$$F(z) = \begin{bmatrix} \frac{0.0007\,z + 0.0007}{z^2 - 1.99\,z + 0.99} & \frac{0.0001\,z + 0.0001}{z^2 - 1.96\,z + 0.96} & \frac{0.0007\,z + 0.0007}{z^2 - 1.92\,z + 0.92} & \frac{0.0003\,z + 0.0003}{z^2 - 1.94\,z + 0.94} \\ \frac{0.0007\,z + 0.0007}{z^2 - 1.99\,z + 0.992} & \frac{0.0001\,z + 0.0001}{z^2 - 1.96\,z + 0.96} & \frac{0.0007\,z + 0.0007}{z^2 - 1.92\,z + 0.92} & \frac{0.0003\,z + 0.0003}{z^2 - 1.94\,z + 0.94} \\ \frac{0.0007\,z + 0.0007}{z^2 - 1.99\,z + 0.99} & \frac{0.0001\,z + 0.0001}{z^2 - 1.96\,z + 0.96} & \frac{0.0007\,z + 0.0007}{z^2 - 1.92\,z + 0.92} & \frac{0.0003\,z + 0.0003}{z^2 - 1.94\,z + 0.94} \end{bmatrix} \tag{46}$$

In order to increase the robustness of system, we add a feedback filter in feedback loop. The expression of the added filter is as follow:

$$G(z) = diag\left\{G_{f11}(z) \quad G_{f22}(z) \quad G_{f33}(z) \quad G_{f44}(z)\right\} \tag{47}$$

where,

$$G_{f11}(z)=\frac{z}{z^2-0.001z+0.004},G_{f22}(z)=\frac{z-0.0001422}{z^2-0.0002846z+0.00000125} \tag{48}$$

$$G_{f33}(z)=\frac{0.9971z-0.006362}{z^2-0.009359z+0.0001379},G_{f44}(z)=\frac{z-0.00008516}{z^2-0.000001784z} \tag{49}$$

The expression of the inverse of $G_m(z)$, $G_m^*(z)$ is given as follows

$$G_m^*(z)=\begin{bmatrix} G_{m11}^*(z) & G_{m12}^*(z) & G_{m13}^*(z) & G_{m14}^*(z) \\ G_{m21}^*(z) & G_{m22}^*(z) & G_{m23}^*(z) & G_{m24}^*(z) \\ G_{m31}^*(z) & G_{m32}^*(z) & G_{m33}^*(z) & G_{m34}^*(z) \end{bmatrix} \tag{50}$$

where,

$$G_{m11}^*=\frac{-0.023z^2-0.0006079z}{z^2-0.0366z+0.0003355}z^{-1},G_{m12}^*=\frac{0.000699z-0.0005}{z^2-0.2707z+0.01832} \tag{51}$$

$$G_{m13}^*=\frac{0.01035z^2+0.000875z}{z^2-0.03663z+0.0003355}z^{-1},G_{m14}^*=\frac{-0.04769z^2-0.006181z}{z^2-0.139z+0.004828}z^{-1} \tag{52}$$

$$G_{m21}^*=\frac{-0.3386\,z^2+0.07526\,z-0.001504}{z^2-0.03663\,z+0.0003355},G_{m22}^*=\frac{-0.7011\,z^2+0.3355\,z-0.04626}{z^2-0.2707\,z+0.01832}z^{-1} \tag{53}$$

$$G_{m23}^*=\frac{0.01847\,z^2+0.001332\,z}{z^2-0.03663\,z+0.0003355}z^{-1},G_{m24}^*=\frac{-0.0025\,z^2-0.03611\,z-0.00231}{z^2-0.139\,z+0.004828} \tag{54}$$

$$G_{m31}^*=\frac{-0.07642\,z^2+0.003053\,z}{z^2-0.03663\,z+0.0003355}z^{-1},G_{m32}^*=\frac{0.105\,z^2-0.05384\,z-0.06541}{z^2-0.2707\,z+0.01832}z^{-1} \tag{55}$$

$$G_{m33}^*=\frac{-1.182e\text{-}08\,z^2-7.598e\text{-}07\,z}{z^2-0.03663\,z+0.0003355},G_{m34}^*=\frac{0.04368\,z^2+0.002231\,z-7.429e\text{-}06}{z^2-0.139\,z+0.004828} \tag{56}$$

The final expression of the internal controller based on SVD described by the equation (26) is the following

$$C_{SDV}(z)=G_m^*(z)G_{m+}(z)W_v F(z)W_v^{-1} \tag{57}$$

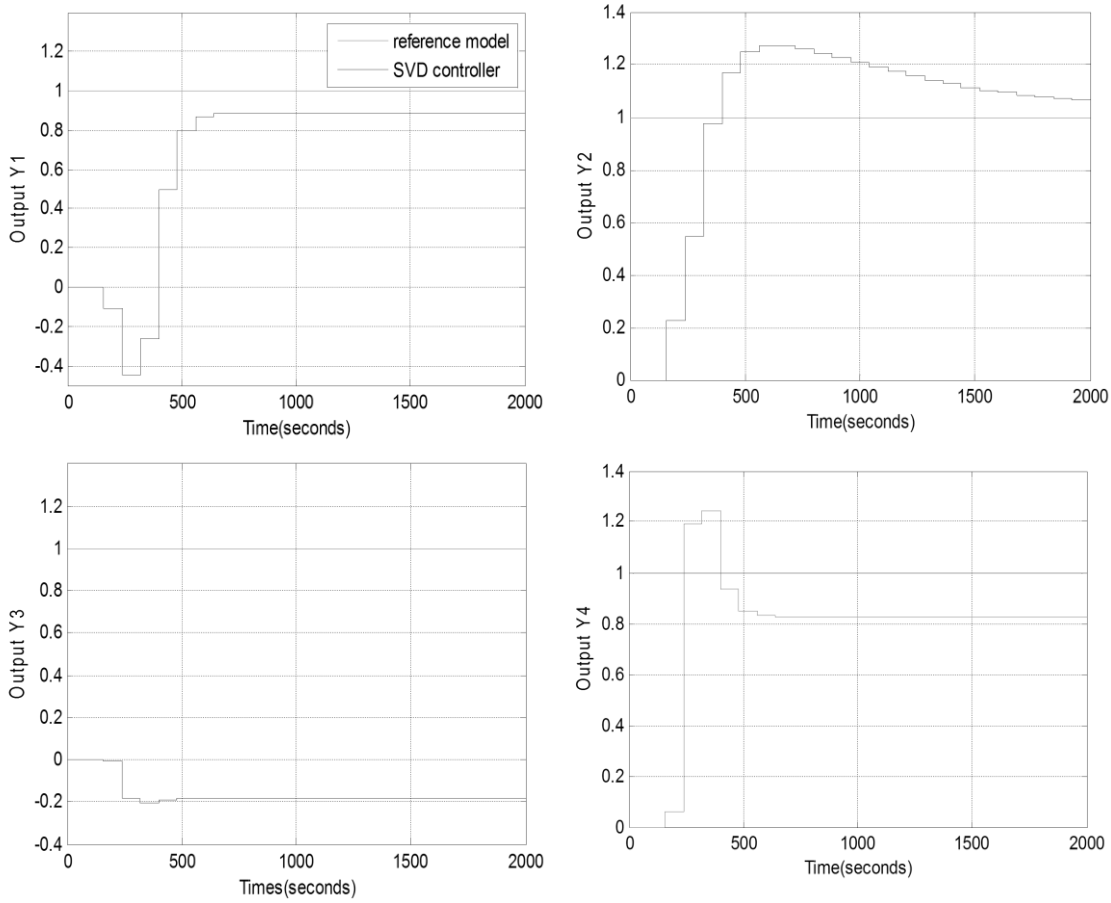

Figure 6. Outputs System with internal controller based on SVD.

SIMULATION RESULTS USING THE VIRTUAL INPUT CONTROLLER

Dealing now with our proposed approach, the Virtual Input methods applied on the same system studied previously. Considering the same underactuated system with three control and four outputs. The system transfer matrix *G(z)* is given by (33).

Let us consider the case of perfect modeling $G(z) = G_m(z)$, The augmented model transfer function is of dimension 4 × 4. The system and the model outputs are expressed respectively by (58) and (59).

$$\begin{bmatrix} y_1 & y_2 & y_3 & y_4 \end{bmatrix}^T = G(z)u \tag{58}$$

$$\begin{bmatrix} y_{m1} & y_{m2} & y_{m3} & y_{m4} \end{bmatrix}^T = \begin{bmatrix} G_{m11} & G_{m12} & G_{m13} & G_{m14} \\ G_{m21} & G_{m22} & G_{m23} & G_{m24} \\ G_{m31} & G_{m32} & G_{m33} & G_{m34} \\ G_{m41} & G_{m42} & G_{m43} & G_{m44} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} \tag{59}$$

The model transfer functions G_{m11} until G_{m43} are successively chosen to be close to G_{11} until G_{43}. The augmenting (n × (n-m)) virtual column $G_{m14}, G_{m24}, G_{m34}$ and G_{m44} are chosen as first order transfer functions so that they ensure the invertibility conditions of the matrix $G_m(z)$.

It is mandatory to study the discrete-time internal controller stability, in order to set the stability interval of the gain matrix K_1 which ensures the stability of the IMC structure. The expression of the controller (29) detailed in the previous section can be reformulated using the representation of sate as:

$$\begin{cases} x(k+1) = F\,x(k) + H\,e(k) \\ u(k) = \mathrm{C}\,x(k) \end{cases} \tag{60}$$

Where $x(k) \in \mathbb{R}^n$, $e(k) \in \mathbb{R}^n$ and $u(k)\ \mathbb{R}^n$, are the state, input and output vectors respectively of the controller, the matrices F, H and C are known constant matrices. The stability condition highlighted by Lyapunov theory allows to assess the necessary and sufficient condition of the controller stability. Using this theory, we can conclude that the system presented by the equation (60) is stable if and only if there exists a positive definite matrix $P = P^T > 0$, satisfying the following Lyapunov inequality [21]:

$$P > 0\ ,\ \left(F^T P F - P\right) < 0 \tag{61}$$

Solving the LMI equation (61), the interval of the gain K_1 which assures the stability of the internal controller is $-0.1 < K_1 < 0.01 \times I_4$. In the case of this system, we choose $a = 0.01$ such that $K_1 = \alpha \times I_4$.

The gain matrix K_2 relative to $K_1 = 0.01 \times I_4$ is given by:

$$K_2 = \begin{bmatrix} 1.1869 & -0.0177 & -0.7142 & 0.5450 \\ 0.9388 & 5.3832 & -5.6314 & 0.3094 \\ 0.6865 & 0.1494 & 0.7212 & -0.5571 \\ 0.0733 & 0.0105 & 0.8240 & 1.1118 \end{bmatrix} \tag{62}$$

The set responses of the studied system with both controller design is presented through the Figure 6 and Figure 7; we note that the steady state error is not null in the case where we used the controller based on SVD.

COMPARISON ANALYSIS

After using both Singular Value Decomposition and Virtual Inputs approaches, it seems to be quite interesting making a comparison between them showing their effectiveness in terms of stability, robustness, precision and tracking signal.

From the data presented in table 1, it is relevant to note that the desired specifications of the closed-loop responses are not met with the SDV controller and this is expected because of the problem of the interaction which are too strange as illustrated by the decoupling of the controller and the simulation results obtained with the SVD controller in Figure 6. SVD method has a problem if we do not make the right choice of the parameter of the filter, in this case, the approximation cannot be made with effectiveness and the system can be diverging. Added to that this approach towards the under-actuated process is to square the system by make the decoupling of the model affecting the characteristic of the system. Unfortunately, this operation can decrease the performance of the system and making it so poor by neglecting some information's.

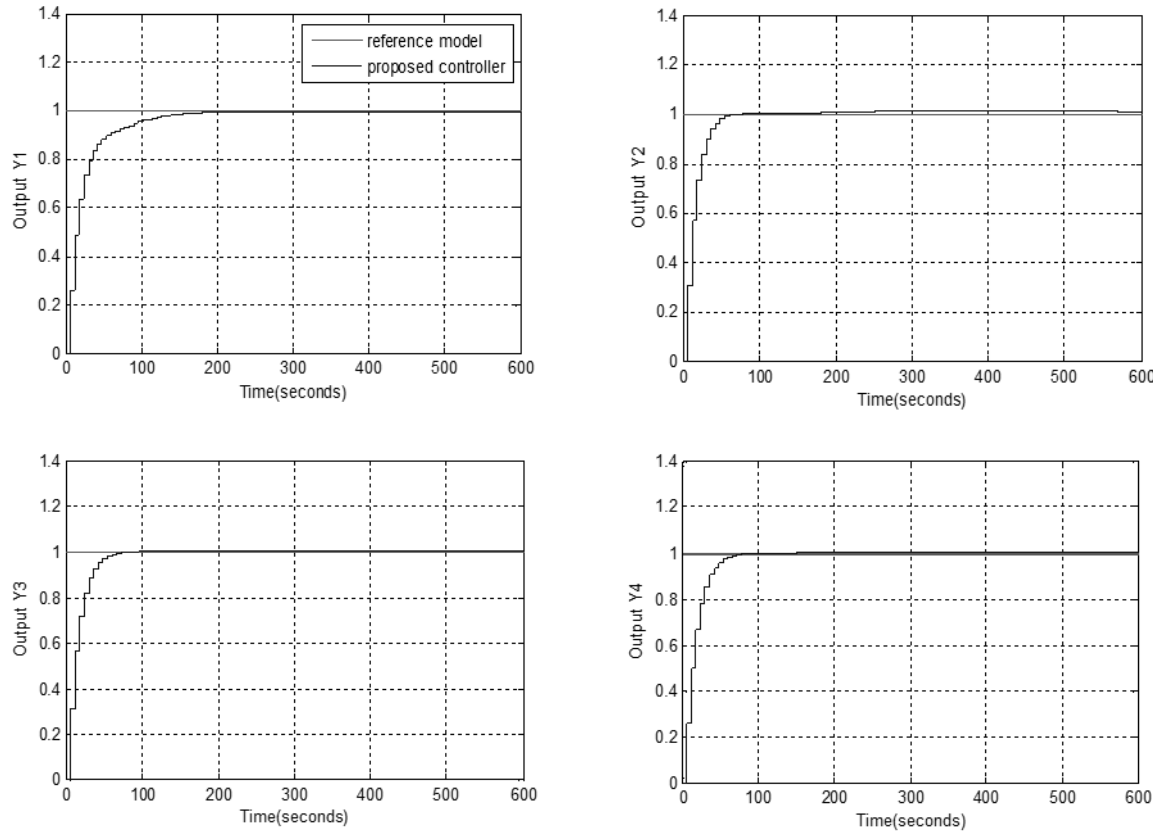

Figure 7. System outputs with internal controller based on Inputs Virtual.

The internal model controller using a Virtual Inputs as inversion approach of the model $G_m(z)$ better performances in terms of rise time, settling time and in terms of error dynamics precision are obtained, as shown by table1, these results are satisfactory because the steady state is zero for all responses as shown in Figure 7. This mean that the set-point tracking is ensured. The internal model control using the Virtual Input approach is having the benefits of small overshoot, faster tracking features, and less interaction compared to the Singular Value Decomposition method as indicated by the Figures 6 and 7 and by the table below.

The Virtual Input method may be fails when the system has a disturbance acting on the outputs of the system. For this reason, with load disturbance the response needs to be controlled to attain robustness and performance.

Table 1. Quantitative comparison between IMC Controller based SVD and Virtual Inputs.

	Parameters to control	**Outputs of System**			
		Y1	**Y2**	**Y3**	**Y4**
IMC controller based on SVD	Rise Time (s)	93	160,6	-6	240
	Steady state error (%)	12,99	1,96	66,24	21,36
	Settling time (s)	560	320	247	400
IMC controller based on VI	Rise Time (s)	53,95	23,92	30,02	30,14
	Steady state error (%)	0	0	0	0
	Settling time (s)	96,02	42,05	42,05	47,97

In order to verify the robustness of the proposed internal model controller, we added a white noise with a variation of 0,2 acting on the outputs of the system in the interval of time 100 s, 150 s, 200 s, 250 s. The influence of this noise is observed in Figure 8.

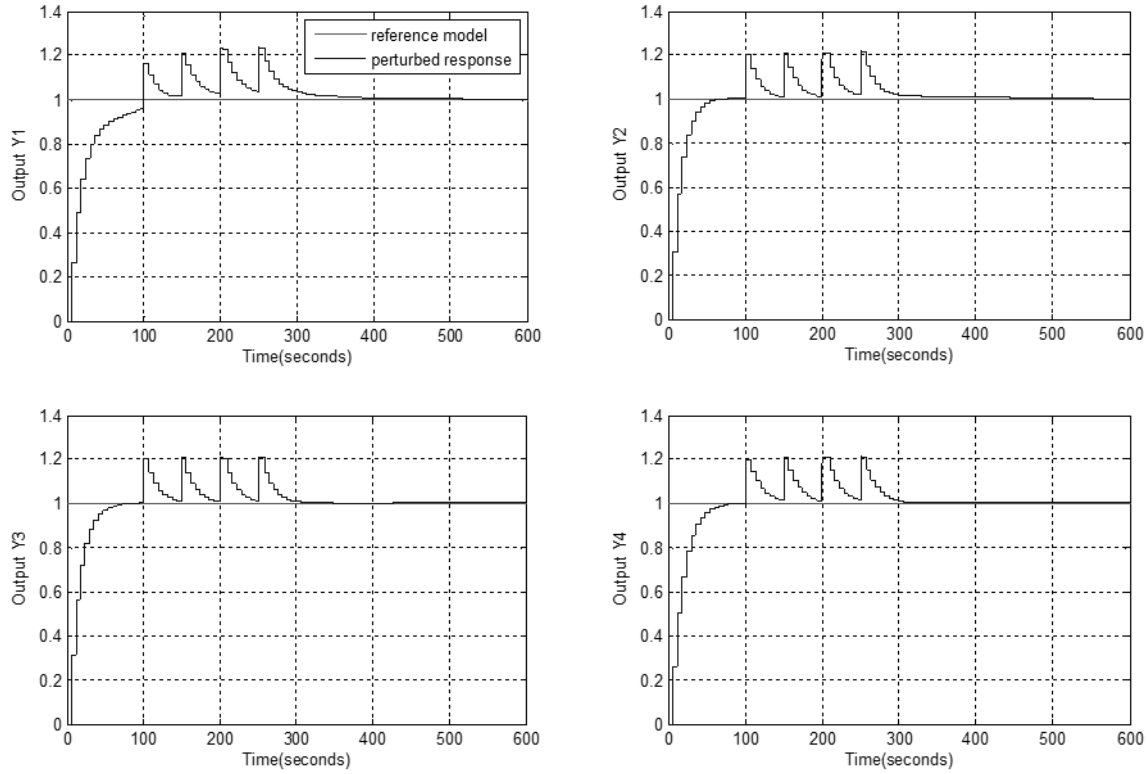

Figure 8. System outputs with disturbed internal controller based on Virtual Inputs.

It is shown that the closed loop system is not sensitive to noise. We conclude that the method it offers strong robustness in case of the given perturbation and good tracking features.

CONCLUSION

To control an under-actuated industrial process demands not only to know all the characteristic of his model, but also to achieve the desired performance when there exist load disturbances. Both Virtual Input and SVD approaches are applied to distillation column process, to prove the controller achievability for this industrial process and that it is the most efficient approach in terms of rapid response, set-point tracking and disturbance rejection. Through a comparative analysis, the Virtual Inputs approach avoids the complex calculation, such as calculate the inverse of the matrix, the controller structure is simple, has a few tuning parameters, and is easy to be accepted by operators.

The simulation results show that our proposed method (Virtual Inputs) ensures suitably the set-point tracking and the disturbance rejection for the external disturbance as illustrated in distillation control problem suggested by Doukas and Lyben and gives better results compared to SVD approach. The modified internal model control scheme gives more degree of freedom in the controller design technique in order to improve the performance of the controlled system.

As future works, it will be interesting to handle with unstable multivariable non-square systems and non-linear systems as proposed in order to design an internal controller with more flexibility and to apply the proposed Virtual Input controller to real processes.

REFERENCES

[1] De Luca, A.; Mattone, R. and Oriolo, G.*: Stabilization of underactuated planar 2R manipulator*.
International Journal of Robust and Nonlinear Control **10**(4), 181-198, 2000,
http://dx.doi.org/10.1002/(SICI)1099-1239(20000415)10:4<181::AID-RNC471>3.0.CO;2-X,

[2] Garcia, C.E. and Morari, M.: *Internal model Control. 2. Design procedure for multivariable systems.*
Industrial & Engineering Chemistry Process Design and Development **24**(2), 472-484, 1985,
http://dx.doi.org/10.1021/i200029a043,

[3] Zhang, X. and Pang, H.: *Novel concise robust control design for non-square systems with multiple time delays*.
Journal of Nature and Science **1**(2), 1-4, 2015,

[4] Jin, Q.; Jiang, B.; Wang, Q. and Shan, G.: *Decoupling internal model control for non-square processes based on equivalent transfer function*.
Transaction of the Institute of Measurment and Control **36**(8), 1114-1131, 2014,
http://dx.doi.org/10.1177/0142331214534110,

[5] Shan, G. and Wang Q.: *The design of disturbances rejection controller for decoupling internal model control system of non-square processes*.
Proceedings of the International Workshop on Computer Science in Sports, 2013,
http://dx.doi.org/10.2991/iwcss-13.2013.41,

[6] Pamela, D.; Jebrajan, T. and Baby, B.: *Real time implementation of internal model controller for temperature process*.
International Conference on Renewable Energy and Sustainable Energy. Coimbatore, pp.162-165, 2013,
http://dx.doi.org/10.1109/ICRESE.2013.6927807,

[7] Bing-Jin, Q.; Zhao, L.; Hao, F. and Wen Liu, S.: *Design of a multivariable internal model controller based on singular value decomposition*.
The Canadian Journal of Chemical Engineering **91**(6), 1103-1114, 2013,
http://dx.doi.org/10.1002/cjce.21735,

[8] Morari, M. and Garcia, C.E.: *Internal model control, A unifying review and some new results.*
Industrial & Engineering Chemistry Process Design and Development **21**(2), 308-323, 1982,
http://dx.doi.org/10.1021/i200017a016,

[9] Touati, N.; Saidi, I.; Dhahri, A. and Soudani, D.: *Internal multimodel control for nonlinear overactuated systems*.
Arabian Journal for Science and Engineering **44**(3), 2369-2377, 2019,
http://dx.doi.org/10.1007/s13369-018-3515-5,

[10] Qi-Bing, J.; Si-Wen, L.; Ling, Q. and Li-Ting, C.: *Internal model control based on singular value decompostion and its application to non-square processes*.
Acta Automatica Sinica **37**(3), 354-359, 2011,
http://dx.doi.org/10.3724/SP.J.1004.2011.00354,

[11] Tian, C.: *Decoupling internal model control method for operation of indutrial process*.
Journal of ACTA Automatica Sinica **35**(10), 1362-1368, 2009,
http://dx.doi.org/10.3724/sp.j.1004.2009.01362,

[12] Penrose, R.: *A generalized inverse for matrices*.
Proceedings of the Cambridge Philosophical Society **51**(3), 406-413, 1955,
http://dx.doi.org/10.1017/S0305004100030401,

[13] Sit Lee, W. and Anderson, B.D.O.: *New filters for internal model control design*.
International Journal of Robust and Nonlinear Control **4**(6), 757-775, 1994,
http://dx.doi.org/10.1002/rnc.4590040605,

[14] Wang, Q.G.; Zhang, Y. and Chiu, M.S.: *Decoupling internal model control for multivariable systems with multiple time delays*.
Chemical Engineering Science **57**(1), 115-124, 2002,
http://dx.doi.org/10.1016/S0009-2509(01)00365-7,

[15] Normey-Rico, J.E.; Bordons, C. and Camacho, E.F.: *Improving the robustness of dead-time compensationg PI controllers*.
Control Engineering Practice **5**(6), 801-810, 1997,
http://dx.doi.org/10.1016/S0967-0661(97)00064-6,

[16] Chen, J.; Zhang, B. and Qi, X.: *A new control method for MIMO first order time delay non-square system*.
Journal of Process Control **21**(4), 538-546, 2011,
http://dx.doi.org/10.1016/j.jprocont.2011.01.007,

[17] Bejaoui, I.; Saidi, I. and Soudani, D.: *Internal model control of MIMO discrete under-actuated systems with real parametric uncertainty*.
Proceedings of the 2nd International Conference on Advanced Systems and electric Technologie, Hammamet, pp.308-314, 2018,
http://dx.doi.org/10.1109/ASET.2018.8379874,

[18] Saidi, I.; Touati, N.; Dhahri, A. and Soudani, D.: *A comparative study on existing and new methods to design internal model controllers for non-square systems*.
Transactions of the Institute of Measurement and Control **41**(13), 3637-3650, 2019,
http://dx.doi.org/10.1177/0142331219834608,

[19] Bejaoui, I.; Saidi, I. and Soudani, D.: *Internal model control of MIMO discrete under-actuated systems with real parametric uncertainty*.
Proceedings of the 2nd International Conference on Advanced Systems and electric Technologie, Hammamet, pp.308-314, 2018,
http://dx.doi.org/10.1109/ASET.2018.8379874,

[20] Bhat-Vinayambika, S.; Shanmuga-Priya, S. and Thirunavukkarasu I.: *A Comparative Study on Control Techniques of Non-square Matrix Distillation Column*.
International Journal of Control Theory and Applications **8**(3), 1129-1136, 2015,

[21] Chang, J.W. and Yu, C.C.: *The relative gain for non-square multivariable systems*.
Chemical Enginnering Science **45**(5), 1309-1323, 1990,
http://dx,doi.org/10.1016/0009-2509(90)87123-A.

Interdisciplinary Description of Complex Systems 19(4), 580-590, 2021

WHICH CHORD PROGRESSIONS SATISFY US THE MOST? THE EFFECT OF EXPECTANCY, MUSIC EDUCATION, AND PITCH HEIGHT

Žiga Mekiš Recek*, Zala Rojs, Laura Šinkovec, Petra Štibelj, Martin Vogrin, Brina Zamrnik and Anka Slana Ozimič

University of Ljubljana, Faculty of Arts, Department of Psychology
Ljubljana, Slovenia

DOI: 10.7906/indecs.19.4.9
Regular article

Received: 25 January 2021.
Accepted: 21 May 2021.

ABSTRACT

Music is an integral part of our everyday lives. Through continuous exposure to a particular music style, an individual implicitly learns the laws of music, including the typical progression of chords that accompany the leading melody. Previous research has shown that the typical chord order in compositions is perceived as expected and satisfying, whereas the violations of the typical chord progressions are perceived as unexpected and unsatisfying. In this paper, we investigated how implicit musical knowledge influences satisfaction during listening to expected and unexpected chord progressions by taking into account the participant's music education and the overall pitch height of the chordal sequences. Ninety-seven participants (43 musicians and 54 non-musicians) took part in the experiment. They were asked to rate the degree of their satisfaction during listening to expected and unexpected chord progressions, either during the high-pitch or low-pitch height conditions. The results showed that the participants were more satisfied with expected than unexpected chord progressions, confirming previous findings on the role of implicit learning of rules of harmony. Although results did not reveal an effect of music education during listening to expected chord progressions, musicians evaluated unexpected progressions as less satisfying than non-musicians, suggesting that musicians' are more susceptible to violations of typical chord order. Finally, the results have shown that the difference in satisfaction between expected and unexpected progressions was larger in high-pitch vs. low-pitch condition, suggesting that under low-pitch condition, chord progressions were more difficult to discriminate, confirming the theory of low-interval limit.

KEY WORDS

harmonic progression, implicit learning, low interval limit, music education, satisfaction

CLASSIFICATION

APA: 2300, 2326, 2340, 2343, 2360

JEL: I39

*Corresponding author, η: ziga.mekisrecek@gmail.com; -;-

INTRODUCTION

Music is an integral part of our everyday lives. Through exposure to a particular music style, an individual implicitly learns the laws of rhythm, melody, harmony, and other aspects of sound organization, forming implicit expectations about how these features are supposed to develop through auditory experience [1]. If these expectations are met, the individual experiences satisfaction while listening to music and vice versa – if the expectations are violated, the listener is not satisfied with the listening experience [2]. The aim of this paper was to investigate how implicit musical knowledge influences satisfaction during listening to expected and unexpected harmonic musical sequences and how this interacts with the participant's music education and the overall pitch height of the sequences.

Music is formed by a series of notes played sequentially or simultaneously, forming a melody and harmony, respectively. A chord, which is the fundamental building block of harmony, consists of three or more simultaneously played notes, while a chord progression is a sequence of chords. The most common chord order in traditional western music is a tonic – tonic in the first inversion – subdominant – dominant – tonic [3]. Specifically, a typical chord progression begins and ends with a tonic chord, whose keynote matches the scale. For example, in a composition that is written in C major, the keynote is C, and the tonic chord consists of C, E, and G. After the first tonic chord, inverted tonic chords, whose function is to lengthen the sense of the first tonic chord, usually follow. An inverted chord simply means that a note other than the keynote is at the bottom. Based on the lowest note from the chord (e.g., E or G in C major), we can determine its inversion. In the first inversion, the lowest note is the third of the triad (E in C major), with the fifth (G in C major) and the keynote (C in C major) stacked above it. After the lengthening of the tonic chord with the inverted tonic chords, the subdominant (e.g., in C major, the subdominant chord consists of F, A, C) and the dominant (e.g., in C major, the dominant chord consists of G, B, D) usually follow. Same as at the start of the progression, the chord progression typically ends with the tonic.

With the continuous exposure to western music, individuals implicitly internalize the structure of the typical chord progression to the extent that the typical form of the harmonic syntax in compositions is perceived as expected and satisfying, whereas the violations of the syntax rules are perceived as unexpected and unsatisfying [2, 4]. Loui and Wessel [5] have demonstrated this effect for both musicians and non-musicians. Although musicians were, in general, more satisfied when listening to musical sequences than non-musicians, the pattern of satisfaction to expected and unexpected chord progressions did not differ between musicians and non-musicians [5]. Similarly, the sensitivity to the violations of the typical harmonic context of both musicians and non-musicians has been shown by reaction times [6, 7] and event-related potentials [8, 9] studies. Together these studies show that both musicians and non-musicians implicitly internalize rules of music harmony and make harmonic expectations, on which they base their response for a musical phrase.

In contrast to Loui and Wessel [5], Dellacherie et al. [10] found differences between musicians and non-musicians in how they respond to expected and unexpected musical sequences. Specifically, they investigated the response to consonant and dissonant musical excerpts. Dissonances, which can be defined as a combination of sounds that do not belong to a particular musical style [11], can - similarly to the unexpected chord resolutions - be treated as tensions in western music, which can induce unexpectancy and unpleasantness during listening to music. Dellacherie et al. [10] found greater negative emotional self-reports and physiological response of musicians to dissonances compared to non-musicians, whereas their response to consonant excerpts did not differ. They explain their findings with the possibility that a negative bias is formed through the experience of music training. They

proposed that because of long-term associative learning, a link between dissonance and negative emotions is formed. Consistent with Dellacherie et al. [10], Pagès-Portabella and Toro [12] reported that during listening to dissonances, early right anterior negativity as measured by EEG was observed in both musicians and non-musicians, indicating automatic processing of harmonic inconsistencies (i.e., violations of musical syntax). However, a larger effect of harmonic inconsistencies was reported for musicians vs. non-musicians, suggesting that music training modulates how different violations of the harmonic context are processed.

Since musical education seems to modulate the response to violations of music syntax, another intriguing question is whether the overall pitch height of chord progression also plays a modulatory role. Pitch height seems to be an important factor that composers consider when creating music. In order to achieve clarity and purity between the intervals, the tones are presented above the so-called 'low interval limit', a concept that determines the lowest pitches at which intervals can still be perceived without sounding uncollected or muddy [13]. When the lower tones become too layered, they create dissonance and lose their clarity [14]. The limits are not absolute but represent areas below which there is a risk that the resultant sound will not work well within a normal harmonic context. To our knowledge, no study so far has investigated how pitch height influences satisfaction while listening to expected and unexpected chord progressions. Based on the low interval limit concept, we hypothesize that resolutions of chord progressions in the low-pitch condition will be experienced as unfocused, which may prevent participants from discriminating between the types of resolutions. This should result in a smaller difference in satisfaction between expected and unexpected chord progressions in low- vs. high-pitch conditions.

In summary, our study aimed to investigate how implicit musical knowledge affects satisfaction while listening to expected and unexpected resolutions of chord progressions and how this interacts with music education and the overall pitch height of chord progressions. Because of the internalization of the typical harmonic syntax [7, 8, 15], we assume that both musicians and non-musicians will experience chord progressions that are frequently used and are typical in western music as more satisfying than uncommon chord progressions. Furthermore, based on Loui and Wessel's [5] study, we expected that the overall satisfaction for both expected and unexpected chord progressions would be higher in musicians vs. non-musicians. As studies investigating the effect of musical education on the perception of violations of musical syntax do not yield consistent findings (e.g. [5, 10, 12]), two alternative hypotheses can be formed. First, we might expect that the effect of expectancy will be larger in musicians vs. non-musicians, as their extensive exposure to music might lead to even stronger internalization of musical syntax [10]. The alternative is that the effect of expectancy will not differ between musicians and non-musicians, which would indicate that both groups implicitly learn the harmony to a similar extent and that this is independent of music education [5]. Lastly, we were interested in the effect of pitch height of the chord progressions. As explained above, based on the low-interval limit concept, we expect the effect of expectancy to be larger in high vs. low-pitch positions, as in lower pitch positions, chords might be perceived as unclear and dissonant [14], making it more difficult to discriminate between expected and unexpected resolutions and thus reducing the effect of expectancy.

METHOD

PARTICIPANTS

194 Slovenian university and high school students, including students of music conservatories, signed the informed consent to participate in an online experiment. The results of 89 participants who did not complete the experimental task were excluded from the analysis.

Further, the results of another 8 participants were excluded from the analysis because they completed the task in less than 10 minutes, which was estimated to be the minimal time needed to perform the task. The final sample included 97 participants (71 females), aged from 15 to 25 years (M = 18,37, SD = 2,07*)*. Those with more than eight years of formal education in music theory (N = 43 participants) were considered musicians, while the participants with less than eight years of formal education in the field of music theory were considered non-musicians (N = 54 participants).

APPARATUS AND MATERIALS

Nine typical chord progressions that consisted of five chords (Table 1, Figure 1) were constructed for the experiment. They either had an expected (the progression resolved on the I. chord or the tonic of the key) or unexpected (the progression resolved on the IV. chord or the subdominant of the key) resolution.

Table 1. List of chord progressions with an expected and an unexpected resolution.

Expected resolution	Unexpected resolution
G-C-F-G-C (V-I-IV-V-I)	G-C-F-G-F (V-I-IV-V-IV)
C-a-F-G-C (I-vi-IV-V-I)	C-a-F-G-F (I-vi-IV-V-IV)
e-a-d-G-C (iii-vi-ii-V-I)	e-a-d-G-F (iii-vi-ii-V-IV)
C-a-d-G-C (I-vi-ii-V-I)	C-a-d-G-F (I-vi-ii-V-IV)
C-G-a-F-C (I-V-vi-IV-I)	C-G-a-F-F (I-V-vi-IV-IV)
C-F-a-G-C (I-IV-vi-V-I)	C-F-a-G-F (I-IV-vi-V-IV)
C-e-F-G-C (I-iii-IV-V-I)	C-e-F-G-F (I-iii-IV-V-IV)
C-F-C-G-C (I-IV-I-V-I)	C-F-C-G-F (I-IV-I-V-IV)
C-F-d-G-C (I-IV-ii-V-I)	C-F-d-G-F (I-IV-ii-V-IV)

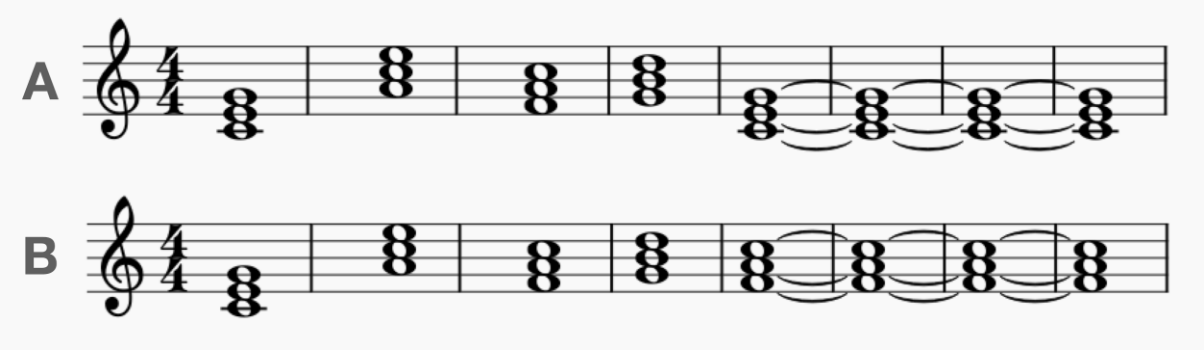

Figure 1. Chord progression examples. A) Chord progression with an expected resolution (the progression ends on the I. chord). B) Chord progression with an unexpected resolution (the progression ends on the IV. chord).

The constructed expected and unexpected chord progressions were recorded in low and high pitch conditions. In low pitch condition, the tonic of the chord progression was note C2 (65,41 Hz), and in high pitch condition, the tonic of the progression was note C4 (262 Hz). Each of the

constructed chord progressions was additionally transposed by or two semitones higher or lower than the original progression, resulting in two starting points for each chord progression (original, transposed).

All progressions were recorded in the time signature of 4/4 in the tempo of 130 beats per minute within eight bars. All of the chords were played within one bar, except for the last, resolving chord, which was played over the last four bars. The progressions were recorded digitally using Logic Pro X, ver. 10.4.4. [16], which is a digital audio workstation. We used a virtual sampled upright piano called 'The Gentleman' [17] that runs in Kontakt 6 player [18] – an industry-standard software for sampling. This provided a clean, clear, and natural sound of a piano with a touch of natural reverb. Psytoolkit software [19, 20] was used to present the chord progressions and collect keyboard-click responses.

TASK

The experimental task consisted of 72 trials (9 chord progressions × 2 resolution types × 2 pitch heights × 2 starting points) presented in random order across participants. Each trial started with the presentation of the chord progression lasting for about 15 s. Following the chord progression, the participant's task was to rate each progression on a scale from 0 (not satisfactory at all) to 100 (most satisfactory) depending on how much subjective satisfaction they experienced while listening to it.

PROCEDURE

We used the snowball sampling method to reach the participants. Participants were informed about the study through different web mediums, such as e-mail, Facebook posts, direct messages, etc. Those interested in participating in the study accessed the study through an experimental website, which contained a detailed description of the study's purpose, terms, and the task itself.

Before signing the informed consent, the participants were instructed to perform the task in a quiet environment using headphones, which allowed them to attend to the task without any interruption and to take a break if they felt fatigued. They were able to test whether both left and right channels of their headphones work properly.

After reading and signing the informed consent, the participants were asked to provide information regarding their age and musical education. After they have read the instructions, they could begin the experimental task. At the end of the experiment, they could provide comments regarding their experience.

RESULTS

To analyze the data, we used a mixed measures three-way analysis of variance (ANOVA) with within-subject factors *type of resolution* (expected vs. unexpected) and *pitch height* (low vs. high) and between-subject factor *music education* (musicians vs. non-musicians).

First, we were interested in the effect of implicit learning of harmony. Specifically, we tested whether chord progressions with expected resolutions satisfy our participants more than chord progressions with unexpected resolutions. The results show that the type of resolution (expected vs. unexpected) significantly affected participants' subjective ratings of satisfaction, $F(1, 95) = 42{,}28$, $p < 0{,}001$, $\eta^2_g = 0{,}051$, revealing that the participants were more satisfied with the expected chord progressions than with the unexpected ones (see Figure 2).

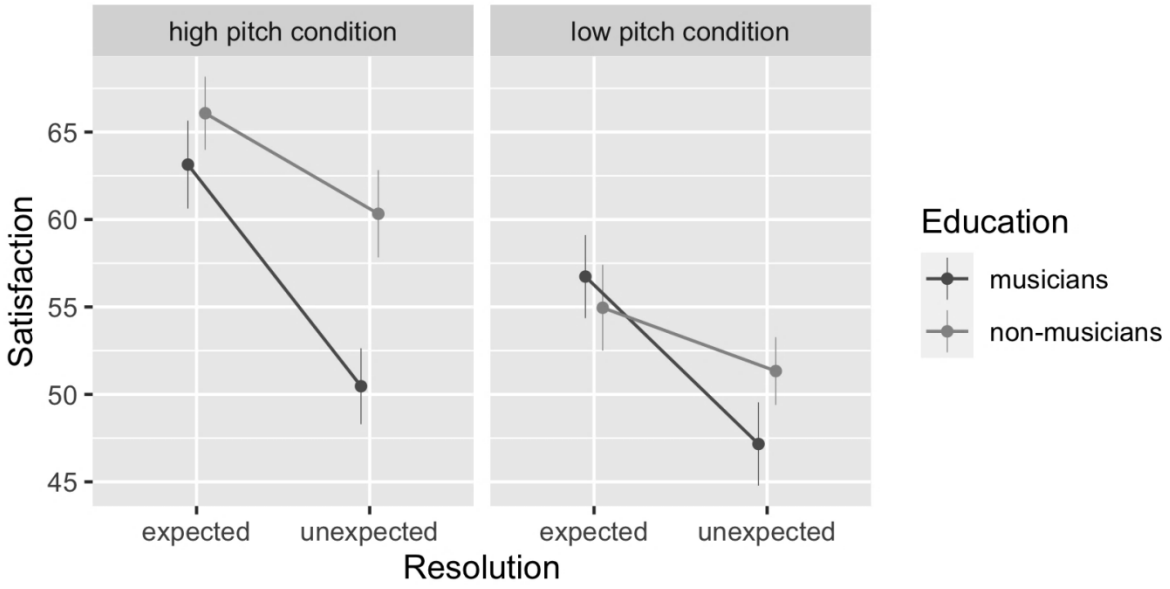

Figure 2. Subjective ratings of satisfaction with chord progressions in musicians and non-musicians depending on the type of resolution of harmonic progression and the progression's pitch condition. Error bars represent 95 % Cousineau-Morey confidence intervals (CI) adjusted so that non-overlap of CIs corresponds to statistically significant difference as calculated with a mixed ANOVA [21].

Next, we were interested in whether musicians are in general more satisfied when listening to chord progressions than non-musicians, regardless of the resolution type. Results did not reveal a main effect of music education (musicians vs. non-musicians) on participants' subjective ratings of satisfaction with chord progressions, $F(1, 95) = 2{,}00$, $p = 0{,}161$, $\eta^2_g = 0{,}013$. However, the results did show a significant interaction between music education and resolution type, $F(1, 95) = 7{,}61$, $p = 0{,}007$, $\eta^2_g = 0{,}054$.

To identify which particular differences between pairs of means are significant, we performed additional *post hoc* tests with FDR-adjusted alpha levels. First, we compared musicians' and non-musicians' satisfaction rates separately for expected and unexpected resolutions. The results did not yield a significant main effect of music education in the expected resolution condition, $F(1, 95) = 0{,}04$, $p = 0{,}841$, $\eta^2_g < 0{,}001$, however, it revealed a significant effect of music education in the unexpected resolution condition, $F(1, 95) = 5{,}59$, $p = 0{,}0270$, $\eta^2 = 0{,}056$, showing that musicians are less satisfied with the unexpected resolutions than non-musicians. Next, we compared satisfaction rates in expected and unexpected conditions separately for musicians and non-musicians. The results revealed a significant effect of resolution in both musicians, $F(1, 42) = 28{,}36$, $p < 0{,}001$, $\eta^2_g = 0{,}096$, and non-musicians $F(1, 53) = 13{,}93$, $p < 0{,}001$, $\eta^2_g = 0{,}041$, reflecting higher satisfaction rates in expected vs. unexpected condition for both musicians and non-musicians.

Lastly, we tested whether the participants are more satisfied with chord progressions in the high pitch condition than in the low pitch condition. The results showed that the progression's pitch height significantly influenced participants' ratings of chord progression satisfaction, $F(1, 95) = 24{,}49$, $p < 0{,}001$, $\eta^2_g = 0{,}010$, suggesting that the participants were more satisfied with the progressions in high pitch condition than with the progressions in low pitch condition. The interaction between the progression's pitch position and type of resolution was also significant, $F(1, 95) = 9{,}37$, $p = 0{,}003$, $\eta^2_g = 0{,}002$, revealing a larger effect of type of resolution in high pitch position than in low pitch position (Figure 2).

We carried out an additional exploratory analysis to test whether participants' satisfaction while listening to chord progressions might have also been affected by the specific sequence of the selected chord progression and whether this might have interacted with the resolution of progressions. We performed a two-way ANOVA with within-subject factors *chord progression* (nine typical chord progressions, see table 1) and *type of resolution* (expected vs. unexpected). The results showed a significant effect of the chord progression on the subjectively experienced satisfaction ratings, $F(8, 768) = 39{,}60$, $p < 0{,}001$, $\eta^2_g = 0{,}021$, as well as significant interaction between type of resolution and type of progression, $F(8, 768) = 3{,}50$, $p < 0{,}001$, $\eta^2_g = 0{,}003$, suggesting that the extent of the effect of resolution type differed between different chord progressions.

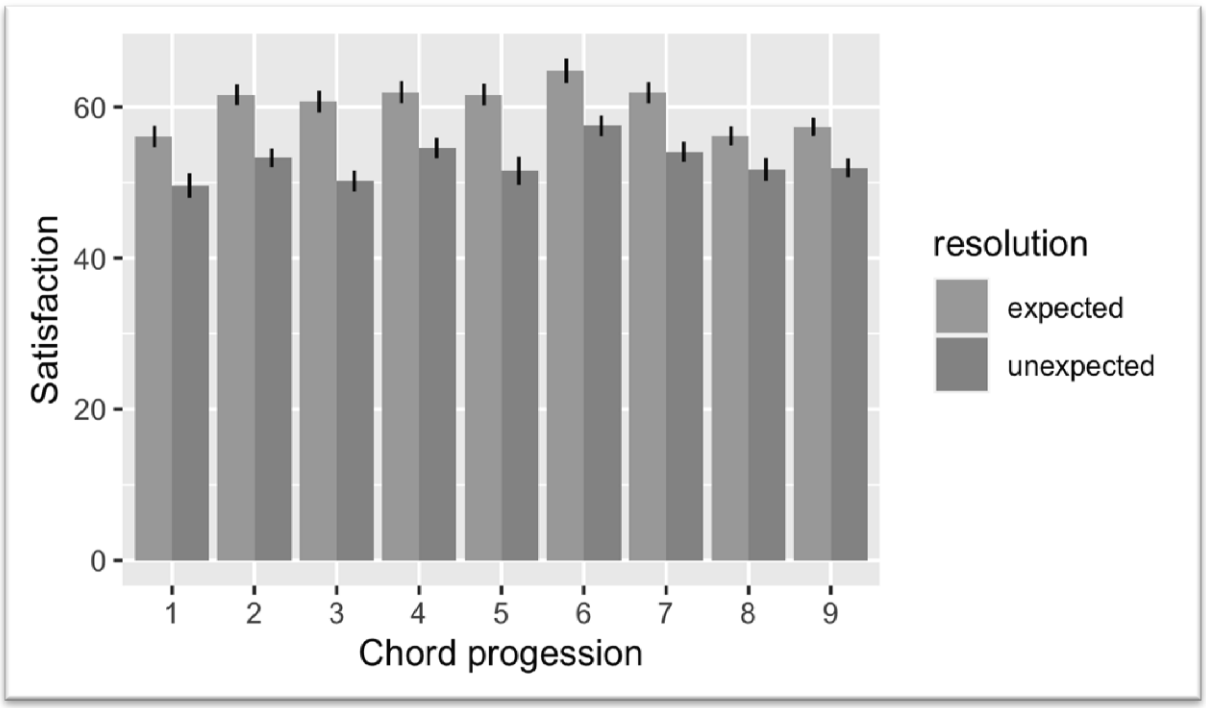

Figure 3. Subjective ratings of satisfaction concerning the nine types of harmonic progressions used (1 to 9) and the type of resolution. Error bars represent 95 % Cousineau-Morey CIs adjusted so that non-overlap of CIs corresponds to statistically significant difference as calculated with a within-subject ANOVA [21].

DISCUSSION

Our study aimed to assess how implicit musical knowledge affects the experience of satisfaction while listening to expected and unexpected chord progressions and how this interacts with music education and the overall pitch height of chord progressions.

Our study's first goal was to evaluate the effect of implicit learning of harmony by comparing participants' satisfaction rates between expected and unexpected chord progressions and to assess whether this effect can be observed in both musicians and non-musicians. The results showed a significant effect of the type of resolution in both groups, reflecting higher satisfaction rates for expected vs. unexpected chord progressions. Our results suggest that both musicians and non-musicians implicitly learned the rules of harmony, which is consistent with previous research (e.g. [7, 8, 15]). Our study confirms that exposure to a particular type of music makes individuals internalize the music style and learn the rules of the harmony implicitly [2].

Next, we checked whether musicians and non-musicians differ in how they experience expected and unexpected resolutions in chord progressions. The satisfaction of musicians generally did not differ from the satisfaction of non-musicians, suggesting that participants, in general, experienced a similar degree of satisfaction while listening to the chord progressions. However, there was a significant interaction between the type of resolution and musical education. Post hoc analyses revealed a difference between musicians and non-musicians in how they experience satisfaction only in the unexpected chord progression condition, showing that the satisfaction in the unexpected resolution condition was smaller in musicians than in non-musicians. This contradicts previous findings indicating that musicians generally felt more pleasure while listening to both expected and unexpected chord progressions (e.g. [5]).

Results allow different interpretations. We can assume that musicians have attained better implicit knowledge based on their musical theory training compared to non-musicians [22]. Implicit knowledge can represent a sound basis for the development of long-term working memory, which enables musicians to recall the appropriate information from long-term memory faster and more efficiently than non-musicians [23]. This includes better recall of the typical chord progressions according to the western music style, as used in our study. Therefore, one explanation of why musicians' level of satisfaction in unexpected chord progressions was smaller than in non-musicians might be related to their superior implicit musical knowledge. The musicians possibly predicted the progressions to end on the tonic, and violations of the typical resolution in unexpected chord progressions had a larger negative effect, reflected in lower satisfaction rates than in non-musicians.

Another possible explanation of the difference in satisfaction of musicians and non-musicians while listening to unexpected resolutions might be related to how we assigned participants into two groups; musicians and non-musicians. In the first group, we included participants who have completed eight or more years of formal music education. That means that they have gained a lot of explicit, theoretical knowledge about western music theory. In the feedback provided after task completion, some musicians reported that they were disrupted because the chord progressions did not end on the tonic but on other scale levels. This suggests that musicians used not only their implicit but also their explicit knowledge to reason and rate their satisfaction level, emphasizing the role of declarative, long-term-explicit knowledge in shaping the experience of satisfaction. The role of implicit and explicit knowledge in shaping musical perception has previously also been recognized by other researchers [24].

Next, the present study results are consistent with the findings of Dellacherie et al. [10], who found that musicians respond more negatively to dissonances than non-musicians, whereas their response to consonances is similar. Dissonances as well as the unexpected resolutions, which were used in our study, can both be treated as tensions in western music, which can induce dissatisfaction during listening to music. Dellacherie et al. [10] suggest that a connection between dissonance and negative emotions is established because of long-term associative learning, possibly leading to a more pronounced effect of expectancy violations in musicians. Our results are also congruent with the findings of the EEG study of Pagès-Portabella and Toro [12], who found a larger early right-anterior negativity (ERAN) in response to listening to dissonance in musicians vs. non-musicians.

In summary, our results show that musical education modulates the processing of violations of typical chord progressions. Our next question was whether the chord progression's overall pitch height might also modulate the satisfaction while listening to expected and unexpected chord progressions. The results showed a significant main effect of pitch height on the reported levels of satisfaction, revealing that the participants rated the resolutions in high pitch height as more satisfactory than the resolutions in low pitch height. This finding is consistent with the low

interval limit theory, which conjectures that the intervals in lower pitch positions sound muddy and unfocussed [14], possibly leading the participants to experience greater dissatisfaction while listening to low vs. high-pitch chord progressions.

Our results also showed a significant interaction between the type of resolution and pitch height, reflecting a larger effect of resolution type in the high-pitch condition. A possible explanation why the experience of listening to resolutions in the low pitch height was less affected by the type of resolution is that under low-pitch condition, resolutions were experienced as unfocused, which prevented the participants from discriminating between the resolution types in the first place. We can conclude that although a low interval limit is not an absolute limit and merely suggests the risk for the chord (or interval) to be experienced as unfocussed, our results show that the pitch height chosen for the low-pitch condition in our study did induce the experience of the dissonance and unclarity to at least some extent. This finding can help understand how the same harmonic pattern can evoke different experiences based on its pitch height.

STUDY LIMITATIONS

The present study has some potential limitations. First, we have noted a significant participant dropout. Of the 194 participants who initially began the test, only 105 completed the entire experiment. In their feedback, many participants reported having found the time necessary to complete the task too long. Additionally, many found the test itself boring. The dropout can therefore be attributed firstly to the duration of the experiment, which took approximately 30 minutes to complete, and secondly, to the type of chord progressions included in the task. The progressions mainly consisted of the same five chords that only varied in the sequence in which they appeared. This repetitiveness is why the participants could have found the task boring. We advise researchers conducting similar studies to use shorter chord presentation times to avoid the experiment being too time-consuming and present the participants with more diverse chord progressions.

Second, the additional exploratory analysis results revealed significant differences in participants' satisfaction between the selected chord progression sequences (i.e., participants preferred certain progressions types over others). Moreover, the effect of specific chord type significantly interacted with the resolution type of chord progressions, suggesting that the extent of the effect of resolution type differed between different chord progressions. This is an important finding that should be taken into account when designing future studies. Future studies should ensure that listeners experience a similar amount of satisfaction while listening to the selected chord progressions so that the studied effect of resolution type would not be confounded with other factors.

In addition to the participant dropout and the types of included progressions, we must also address the suitability of the criteria used to separate the participants into groups of musicians and non-musicians. Only those who reported having obtained at least eight years of formal education in music theory were assigned to a group of musically educated individuals, which resulted in a severely heterogeneous group of non-musicians. The latter included musically completely uneducated participants as well as individuals who have received several years of formal musical education. Despite not having been enrolled in as many years of music theory training as musicians, non-musicians could have had a comparable level of musical knowledge and experience as individuals with eight or more years of formal musical education. In this case, it would be more appropriate to include these individuals in the group of musicians. As the heterogeneity of the participants in the group of non-musicians could have had a potentially significant effect on the obtained results, we advise researchers conducting future studies to define their criteria for distinguishing the musically educated from the musically uneducated participants in a more elaborate and detailed way. We also suggest using music ability tests to objectively determine the participants' music perception skills, such as pitch discrimination.

CONCLUSIONS

In the present study, we explored the effect of expectancy, music education, and pitch height on the satisfaction rates during listening to different chord progressions. As expected, the results showed that the participants were more satisfied with expected than unexpected chord progressions, confirming previous findings on the role of implicit learning of rules of harmony. Although results did not reveal an effect of music education during listening to expected chord progressions, musicians evaluated unexpected progressions as less satisfying than non-musicians, suggesting that musicians process music differently than non-musicians, and are more susceptible to violations of typical chord order. The results have also revealed that the difference in satisfaction between expected and unexpected progressions was larger in high-pitch vs. low-pitch condition, suggesting that chord progressions were more difficult to discriminate under low-pitch condition supporting the theory of low-interval limit. Though the results clearly show that the effect of resolution type is modulated by music education and the overall pitch height of music sequences, the findings need to be interpreted with caution due to methodological limitations.

REFERENCES

[1] Bharucha, J.J.: *Event hierarchies, tonal hierarchies, and assimilation: A reply to Deutsch and Dowling*.
Journal of Experimental Psychology: General **113**(3), 421-425, 1984,
http://dx.doi.org/10.1037/0096-3445.113.3.421,

[2] Salimpoor, V.N.; Zald, D.H.; Zatorre, R.J.; Dagher, A. and McIntosh, A.R.: *Predictions and the brain: how musical sounds become rewarding.*
Trends in Cognitive Sciences **19**, 86-91, 2015,
http://dx.doi.org/10.1016/j.tics.2014.12.001,

[3] Piston, W.; Devoto, M. and Jannery, A.: *Harmony*. 3rd edition.
W.W. Norton and Co., New York, 1987,

[4] Janata, P.: *ERP measures assay the degree of expectancy violation of harmonic contexts in music.*
Journal of Cognitive Neuroscience **7**(2), 153-164, 1995,
http://dx.doi.org/10.1162/jocn.1995.7.2.153,

[5] Loui, P. and Wessel, D.: *Harmonic expectation and affect in Western music: Effects of attention and training.*
Perception and Psychophysics **69**, 1084-1092, 2007,
http://dx.doi.org/10.3758/BF03193946,

[6] Bigand, E. and Pineau, M.: *Global context effects on musical expectancy.*
Perception and Psychophysics **59**, 1098-1107, 1997,
http://dx.doi.org/10.3758/BF03205524,

[7] Bigand, E.; Madurell, F.; Tillmann, B. and Pineau, M.: *Effect of global structure and temporal organization on chord processing.*
Journal of Experimental Psychology: Human Perception and Performance **25**(1), 184-197, 1999,
http://dx.doi.org/10.1037/0096-1523.25.1.184,

[8] Koelsch, S.; Gunter, T.C.; Friederici, A.D. and Schröger, E.: *Brain indices of music processing: 'Nonmusicians' are musical*.
Journal of Cognitive Neuroscience **12**, 520-541, 2000,
http://dx.doi.org/10.1162/089892900562183,

[9] Regnault, P.; Bigand, E. and Besson, M.: *Different brain mechanisms mediate sensitivity to sensory consonance and harmonic context: evidence from auditory event-related brain potentials.*
Journal of Cognitive Neuroscience **13**(2), 241-255, 2001,
http://dx.doi.org/10.1162/089892901564298,

[10] Dellacherie, D.; Roy, M.; Hugueville, L.; Peretz, I. and Samson, S.: *The effect of musical experience on emotional self-reports and psychophysiological responses to dissonance.* Psychophysiology **48**(3), 337-349, 2011, http://dx.doi.org/10.1111/j.1469-8986.2010.01075.x,

[11] Parncutt, R. and Hair, G.: *Consonance and dissonance in theory and psychology: Disentangling dissonant dichotomies.* Journal of Interdisciplinary Music Studies **5**(2), 119-166, 2011, http://dx.doi.org/10.4407/jims.2011.11.002,

[12] Pagès-Portabella, C. and Toro, J.M.: *Dissonant endings of chord progressions elicit a larger ERAN than ambiguous endings in musicians.* Psychophysiology **57**(2), 2020, http://dx.doi.org/10.1111/psyp.13476,

[13] Corzine, V.: *Arranging music for the real world: classical and commercial aspects*. Mel Bay, St. Louis, 2002,

[14] Hoffmann, R.: *Low interval limits*. http://www.robin-hoffmann.com/dfsb/low-interval-limits, accessed June 5th 2020,

[15] Tillmann, B.; Bharucha, J.J. and Bigand, E.: *Implicit learning of tonality: A self-organizing approach*. Psychological Review **107**(4), 885-913, 2000, http://dx.doi.org/10.1037/0033-295X.107.4.885,

[16] Apple Inc.: *Apple Inc. Logic Pro X, ver. 10.4.4*. http://apps.apple.com/us/app/logic-pro/id634148309?mt=12, accessed 12th June 2020,

[17] Native instruments: *The Gentleman*. http://www.native-instruments.com/en/products/komplete/keys/the-gentleman, accessed 12th June 2020,

[18] Native Instruments: *Kontakt 6 Player, ver. 6.2.1*. http://www.native-instruments.com/en/products/komplete/samplers/kontakt-6-player, accessed 12th June 2020,

[19] Stoet, G.: *PsyToolkit - A software package for programming psychological experiments using Linux.* Behavior Research Methods **42**(4), 1096-1104, 2010, http://dx.doi.org/10.3758/BRM.42.4.1096,

[20] Stoet, G.: *PsyToolkit – a novel web-based method for running online questionnaires and reaction-time experiments*. Teaching of Psychology **44**(1), 24-31, 2017, http://dx.doi.org/10.1177/0098628316677643,

[21] Baguley, T.: *Calculating and graphing within-subject confidence intervals for ANOVA.* Behavior Research Methods **44**(1), 158-175, 2012, http://dx.doi.org/10.3758/s13428-011-0123-7,

[22] Thorpe, L.; Cousins, M. and Bramwell, R.: *Implicit knowledge and memory for musical stimuli in musicians and non-musicians.* Psychology of Music **48**(6), 836-845, 2020, http://dx.doi.org/10.1177/0305735619833456,

[23] Ericsson, K.A. and Kintsch, W.: *Long-term working memory.* Psychological Review **102**(2), 211-245, 1995, http://dx.doi.org/10.1037/0033-295X.102.2.211,

[24] Scherer, K.R. and Coutinho, E.: *How music creates emotion: A multifactorial process approach*. In: Cochrane, T.; Fantini, B. and Scherer, K.R., eds.: *Series in affective science. The emotional power of music: Multidisciplinary perspectives on musical arousal, expression, and social control*. Oxford University Press, Oxford, pp.121-145, 2013, http://dx.doi.org/10.1093/acprof:oso/9780199654888.003.0010.

MANUSCRIPT PREPARATION GUIDELINES

Manuscript sent should contain these elements in the following order: title, name(s) and surname(s) of author(s), affiliation(s), summary, key words, classification, manuscript text, references. Sections acknowledgments and remarks are optional. If present, position them right before the references.

ABSTRACT Concisely and clearly written, approx. 250 words.

KEY WORDS Not more than 5 key words, as accurate and precise as possible.

CLASSIFICATION Suggest at least one classification using documented schemes, e.g., ACM, APA, JEL.

TEXT Write using UK spelling of English. Preferred file format is Microsoft Word. Provide manuscripts in grey tone. For online version, manuscripts with coloured textual and graphic material are admissible. Consult editors for details.

Use Arial font for titles: 14pt bold capital letters for titles of sections, 12pt bold capitals for titles of subsections and 12pt bold letters for those of sub-subsections. Include 12pt space before these titles.

Include figures and tables in the preferred position in text. Alternatively, put them in different locations, but state where a particular figure or table should be included. Enumerate them separately using Arabic numerals, strictly following the order they are introduced in the text. Reference figures and tables completely, e.g., "as is shown in Figure 1, *y* depends on x …", or in shortened form using parentheses, e.g., "the *y* dependence on *x* shows (Fig. 1) that…", or "… shows (Figs. 1-3) that …".

Enumerate formulas consecutively using Arabic numerals. In text, refer to a formula by noting its number in parentheses, e.g. expression (1). Use regular font to write names of functions, particular symbols and indices (i.e. sin and not *sin*, differential as d not as *d*, imaginary unit as i and not as *i*, base of natural logarithms as e and not as *e*, x_{n} and not x_n). Use italics for symbols introduced, e.g. $f(x)$. Use brackets and parentheses, e.g. {[( )]}. Use bold letters for vectors and matrices. Put 3pt of space above and below the formulas.

Symbols, abbreviations and other notation that requires explanation should be described in the text, close to the place of first use. Avoid separate lists for that purpose.

Denote footnotes in the text by using Arabic numerals as superscripts. Provide their description in separate section after the concluding section.

References are listed at the end of the article in order of appearance in the text, in formats described below. Data for printed and electronic references is required. Quote references using brackets, e.g. [1], and include multiple references in a single bracket, e.g. [1-3], or [1, 3]. If a part of the reference is used, separate it with semi-colon, e.g. [3; p.4], [3; pp.4-8], [3; p.4, 5; Ch.3]. Mention all authors if there are not more than five of them, starting with surname, and followed with initial(s), as shown below. In other cases mention only the first author and refer to others using et al. If there are two or more authors, separate the last one with the word "and"; for other separations use semicolon. Indicate the titles of all articles, books and other material in italics. Indicate if language is not English. For other data use 11pt font. If both printed version and the Internet source exist, mention them in separate lines. For printed journal articles include journal title, volume, issue (in parentheses), starting and ending page, and year of publication. For other materials include all data enabling one to locate the source. Use the following forms:

[1] Surname, Initial1.Initial2.; Surname, Initial1.Initial2. and Surname, Initial1.Initial2.: *Article title.* Journal name **Vol**(issue), from-to, year, http://www.address, accessed date,

[2] Surname, Initial1.Initial2. and Surname, Initial1.Initial2.: *Book title*. Publisher, city, year,

[3] Surname, Initial1.Initial2.; Surname, Initial1.Initial2., eds.: *Title*. In: editor(s) listed similarly as authors, ed(s).: *Proceedings title*. Publisher, city, year.

If possible, utilise the template available from the INDECS web page.

CORRESPONDENCE Write the corresponding author's **e**-mail address, **t**elephone and **a**ddress (i.e., *η*).

ISSN 1334-4684 (printed)
http://indecs.eu