IJAPR

International Journal of Advances in Production Research

International Journal of Advances in Production Research

Vol. 1 Issue 1, 2024

“Dedicated to the memory of
Prof. Dr. Güneş Gençyılmaz."

The Turkish Society for Production Research

The International Journal of Advances in Production Research (IJAPR) is a scientific and blind peer-reviewed journal published twice a year. It is published by **Üretim Araştırmaları Derneği** (Society for Production Research) utilizing the Dergipark infrastructure in editorial processes.

The journal accepts original and scientific articles written in English. Scientific articles containing original research in all areas of industrial engineering and especially in production and manufacturing systems and advancement in production technologies are included in the journal. Articles sent to the journal must not have been previously published in other journals or accepted for publication.

Blind-review process consisting of two reviewers is applied in the evaluation of the articles. The journal does not charge any fees for accepted articles. DOI numbers are assigned to the articles published in the journal.

The journal aims to facilitate the sharing of scientific advancements achieved in various fields of Industrial Engineering, especially in production and manufacturing systems and advancement in production technologies among esteemed scientists, researchers, and individuals interested in the aforementioned areas through a reputable and peer-reviewed journal publication approach.

The journal's specific areas of focus in the fields of production research, smart factories and manufacturing include:

- Artificial Intelligent Applications
- loT and Production Systems
- Big Data and Analytics
- Lean Production
- Clustering
- Logistics Management
- Intelligent Manufacturing, precision metrology
- New Era in Production Management
- Precision Engineering, Instrumentation, Measurement
- Nanotechnology, Nanometrology and Microengineering
- Decision Making
- New Era in Production Management Education
- Demand Management
- Operations Research Applications
- Energy Management
- Optimization
- Quality Management
- Engineering Management
- Engineering Economy

- Process Management
- Enterprise Resource Planning
- Production Economy
- Project Management
- Production Engineering and Smart Manufacturing
- Facility and Capacity Management
- Production Planning and Control
- Flexible Manufacturing Systems
- Productivity and Performance Management
- Fuzzy Logic
- Healthcare Systems and Management
- Industrial Applications
- Risk Analysis and Management
- Industrial Automation
- Service Systems Management
- Industrial Robotics and Industrial Engineering
- Simulation and Modelling
- Industry 4.0, Industry 5.0 Applications
- Industrial Digitalization
- Machine Learning
- Twin Manufacturing
- Green and sustainable manufacturing
- Multi Criteria Decision Making Applications
- New trends in the Industrial Engineering Education

The scope of the journal includes the publishing of scientific outcomes from diverse domains within Industrial Engineering, especially, Production Research, as well as Advances in Production Technologies. These outcomes undergo peer evaluation before being published. The journal is released twice annually and encompasses the dissemination of research and review articles in the specified fields.

IJAPR is an international electronic journal of the production research community which provides open access to all its content and uses only free programs.

Turkish Society for Production Research, 2024

Contents

International Journal of Advances in Production Research

The Bees Algorithm and Its Applications in Production and Manufacturing

D T Pham[1,a*], Yanjie Song[2,b], Yuanjun Laili[3,c], Asrul Ismail[4,d], Natalia Hartono[1,e], Michael Packianather[5,f] and Marco Castellani[1,g]

The University of Birmingham[1], United Kingdom
School of Reliability and System Engineering, Beihang University[2], China
Beihang University[3], China
University of Portsmouth[4], United Kingdom
Cardiff University[5], United Kingdom
d.t.pham@bham.ac.uk[a], yanjiesong96@gmail.com[b], lailiyuanjun@buaa.edu.cn[c],
asrul.ismail@port.ac.uk[d], n.hartono@bham.ac.uk[e], packianatherms@cf.ac.uk[f],
m.castellani@bham.ac.uk[g]

Abstract— **With the advent of the Fourth Industrial Revolution, production and manufacturing processes and systems have become more complex. Obtaining the best performance from them requires efficient and effective optimisation techniques that do not depend on the availability of process or system models. Such models are usually either not obtainable or mathematically intractable due to the high degrees of nonlinearities and uncertainties in the processes and systems to be represented. The Bees Algorithm is a powerful swarm-based intelligent optimisation metaheuristic inspired by the foraging behaviour of honeybees. The algorithm is conceptually elegant and extremely easy to apply. It has attracted users from virtually all fields of engineering and natural, physical, medical and social sciences. This article reviews the original Bees Algorithm and some of its recent enhancements and gives examples of its applications to optimisation problems in production and manufacturing. The aim is to demonstrate the simplicity, effectiveness and versatility of the algorithm and encourage its further adoption by engineers and researchers across the world to realise smart and sustainable manufacturing and production in the age of Industry 4.0 and beyond.**

Keywords— **Bees Algorithm; swarm intelligence; intelligent optimisation; metaheuristics; production optimisation; manufacturing optimisation; smart manufacturing; Industry 4.0**

I. INTRODUCTION

Today's smart production systems are often distributed and highly interconnected. Many newly developed manufacturing processes are complex, nonlinear and not fully characterised or understood. Obtaining the best performance from those processes and systems requires a different approach from those based on mathematical modelling and optimisation. This is because it is difficult and perhaps even impossible to obtain accurate models for them due to inherent uncertainties. To optimise modern production and manufacturing systems and processes, efficient techniques that do not involve mathematical modelling may be necessary.

The Bees Algorithm, a metaheuristic inspired by the foraging behaviour of honey bees, is eminently suitable for handling complex optimisation problems. All it needs to solve an optimisation problem is a means to evaluate the quality of potential solutions. This is true even if the problem is non-deterministic polynomial complete (NP-complete), i.e., the toughest decision problem in the NP category of problems whose solutions cannot be guaranteed to be found, but can be verified, in polynomial time. NP-complete problems are at least as hard as any other NP problem, meaning finding an efficient algorithm for one solves them all. This is the reason why, since the algorithm was first published by Pham et al. in 2005 [1], it has attracted users from virtually all fields of engineering and natural, physical, medical and social sciences.

This article gives an overview of the Bees Algorithm and its applications in production and manufacturing. Following an introduction to the main ideas underpinning the algorithm, including its global search (exploration) and local search (exploitation) techniques, the article will present recent results and developments relating to the algorithm and its application to production and manufacturing optimisation problems. The article will demonstrate the simplicity, effectiveness and versatility of the algorithm. It is hoped that this will encourage further adoption of this tool to realise smart and sustainable manufacturing and production in the era of the fourth industrial revolution and beyond.

II. BEE FORAGING BEHAVIOUR

---

*Corresponding Author

When searching for food, bees perform both exploration and exploitation. A honey bee colony can have thousands of bees. About 10% of the colony is sent up to 10 km around the hive to go looking for food – those are known as scout bees. Scout bees randomly explore the fields surrounding the hive in search of pollen and nectar. When they have found an interesting food source, they return to the hive to recruit more bees to follow them and help gather the food and bring it back to the colony. To recruit additional bees, the scouts perform a dance known as the waggle dance. The other bees in the hive observe the dance and then decide which scout to follow, i.e., which food source to visit and exploit. According to their fitness, food sources (flower patches) may be visited by more or fewer bees or may be abandoned.

Thus, we have seen the essential points of the Bees Algorithm. Scout bees explore the landscape. Their aim is to find the most abundant source of food to feed the colony. The scouts recruit additional forager bees to help them exploit the discovered food sources. The more abundant the food source, the more foragers will be recruited. When nectar and pollen in a flower patch run short, the bees stop foraging it and move elsewhere. The process continues until the best food source is located.

III. THE ORIGINAL BEES ALGORITHM

Figure 1 is the flow chart of the original Bees Algorithm proposed in 2005 [1, 2].

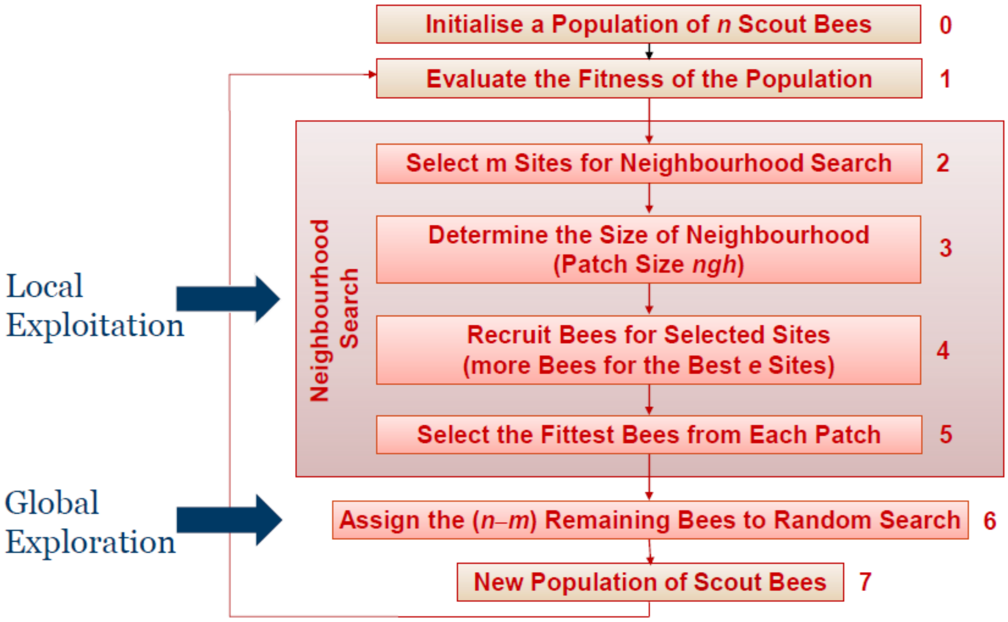


FIGURE 1. THE ORIGINAL BEES ALGORITHM (PHAM ET AL, 2005; PHAM ET AL., 2006)

The algorithm begins with n scout bees moving randomly in space looking for food (Step 0).

The quality of the flower patches (or sites) found by the scout bees is evaluated (Step 1).

A certain number of high-ranking patches (m) are then selected for further exploitation (Step 2).

The size of the search area around those patches (ngh) is defined (Step 3).

The scout bees that found high-quality flower patches are allocated follower bees to help them exploit those patches and their surroundings. The e scout bees that found the very best quality patches are allocated the largest number of followers (Step 4).

In each search area, the bee that found the best flower patch is identified (Step 5). This results in m bees for the next iteration.

Those n- m scouts without followers are again assigned to explore the search space and look for other flower patches (Step 6).

A new population of scouts is thus formed comprising the m bees of Step 5 and the (n-m) bees of Step 6 (Step 7).

The cycle begins again in Step 1 until a stopping criterion is met. That could be when a sufficiently high-quality flower patch has been found or a preset number of iterations has been reached.

The example in Figure 2 illustrates how the algorithm finds the maximum point of a one-dimensional function.

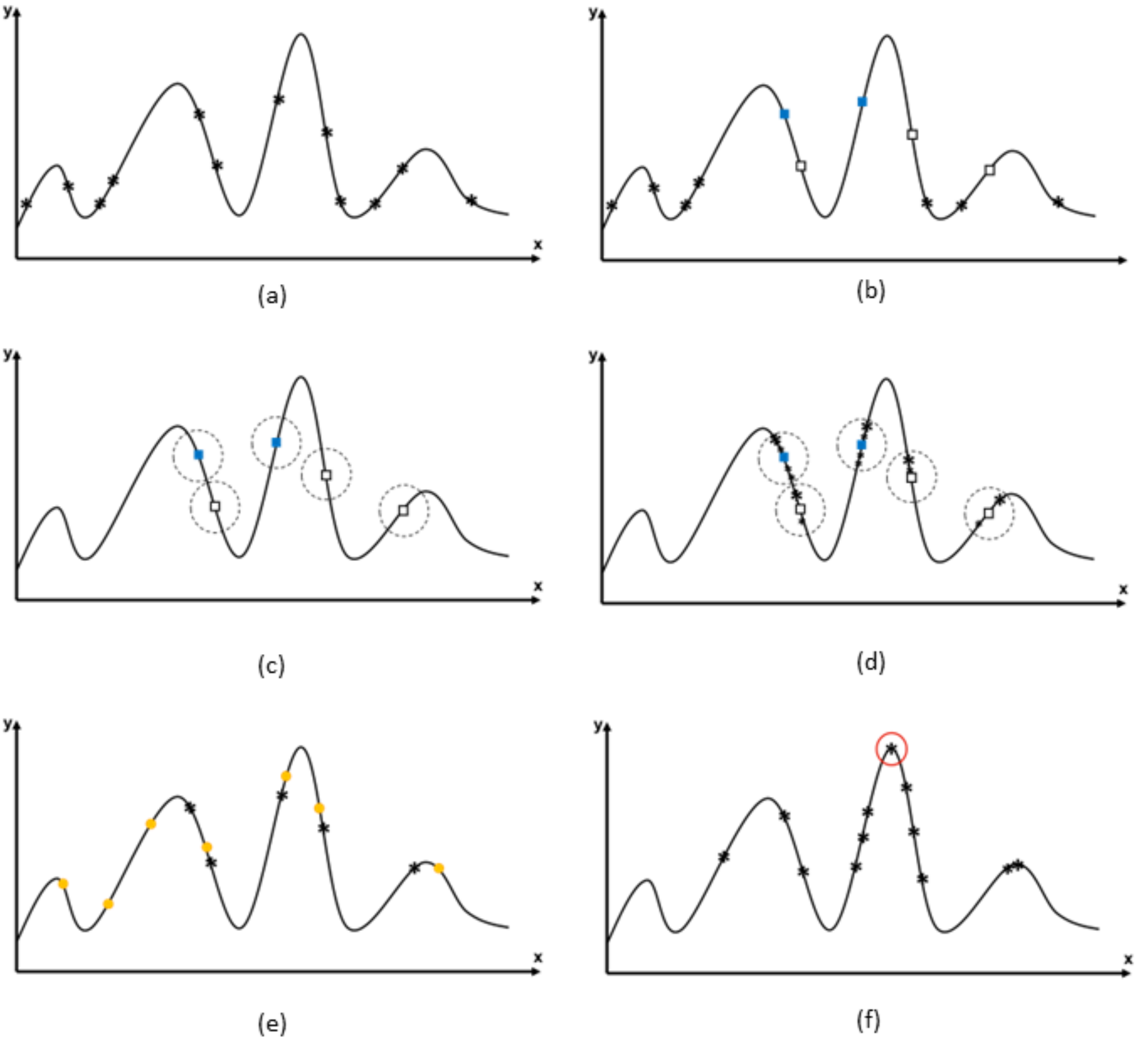


Figure 2. One-D optimisation example (a) Twelve scout bees (n= 12) representing twelve random solutions (asterisks); (b) The top five solutions (m= 5; white and blue squares); (c) Neighbourhoods of the top five solutions; (d) Recruited bees assigned to the neighbourhoods of the top solutions (more recruited bees for the neighbourhoods of the top 2 solutions (e = 2), i.e., the blue squares); (e) The twelve scout bees forming the population in the new iteration (the orange circles represent the solutions found by the 7 scout bees (n-m = 7) not previously used for local exploitation; (f) Convergence to the global optimum (asterisk in the red circle)

IV. Features of the Original Bees Algorithm and Some (Mainly) Early Enhancements

The simplicity of the original Bees Algorithm arises from the fact that it does not require any special interaction mechanism for exchanging information between the bees. For instance, there is no pheromone deposition as with the Ant Algorithm or particle velocity computation like in the case of the Particle Swarm Optimisation algorithm. The Bees Algorithm also has built-in strategies for mobilising resources to exploit promising areas of the search space to locate potential solutions and avoid local optimum traps. The selective recruitment method explained above, whereby foragers focus on the best flower patches, is what makes the algorithm efficient at finding locally optimum solutions.

Although the Bees Algorithm's greedy exploitation method enables it quickly to discover potential solutions, there is a risk of premature convergence to false optima. In addition to random exploration by scout bees, which helps maintain diversity in the solution population to prevent premature convergence, the Bees Algorithm also employs neighbourhood shrinking and site abandonment to escape from local optimum traps. This strategy involves gradually reducing the size of the local search area when the solution does not improve and abandoning that area after a predetermined number of attempts.

In the early years of the development of the Bees Algorithm, it was realised that the new solutions discovered by the scout bees in Steps 7/8 (Figure 1) tended to be poor. Forcing them to compete immediately with those already in the population is counterproductive. However, as a new promising solution may be at the foot of a tall mountain and be worse than old solutions that have stalled at a lower peak, it should be protected and allowed to develop its potential to reach greater heights.

A new category of bees was thus introduced - the "young bees", that is, scout bees that were created in recent iterations of the optimisation process [3]. These new bees are shielded from the stronger adults in the general population. They are permitted to compete for survival only among themselves until they reach "adulthood" after having evolved a given number of iterations when they are ready to ‘fight’ against other more established bees.

The “young bees” idea was a rare excursion into the global search mechanism of the Bees Algorithm. Early work on enhancing the Bees Algorithm tended to focus on local search methods and produced a variety of operators to improve their effectiveness. In addition to mutation and crossover, there were operators such as interpolation, extrapolation and creep [3], as well as those based on TRIZ [4] and Kalman filtering [5]. The use of different operators gave rise to different versions of the Bees Algorithm with different features making them suitable for a variety of applications.

Over the years, several versions of the Bees Algorithm have been developed by hybridising the bee foraging metaphor with other metaheuristics. The Bees Algorithm has also been used in conjunction with computational intelligence tools such as neural networks and fuzzy logic. Table I is a non-exclusive list of hybrid algorithms and paired applications. Many more hybrid versions could be invented by combining the Bees Algorithm with new optimisation methods when available. Similarly, the Bees Algorithm could be synergistically used with other new intelligent tools to tune them, refine their outputs or improve its own performance.

For an extensive survey of other versions of the Bees Algorithm created before 2017, the reader is referred to the article by Hussein et al [6]. A more recent review can be found in the doctoral thesis by Ismail [7]. It will be seen that large amounts of effort were spent on making the algorithm faster. Research was also directed at reducing its parameters from the six or seven that need to be set (namely, n scout bees, m high-performing bees, e ‘elite’ bees, ne forager bees assigned to the ‘elite’ bees, np forager bees assigned to the m-e high-performing bees, ngh, the neighbourhood size, and r, the neighbourhood shrinking rate if neighbourhood shrinking is adopted).

TABLE I
HYBRIDISATION AND COMBINATION

‘Partner’ Algorithm	Method	Authors
Ant Colony Optimisation (ACO)	Using pheromone for increased social interaction.	[8] Packianather *et al*., 2009
Particle Swarm Optimisation (PSO)	‘Elite’ and other successful scouts acting together as a swarm or as leaders of their own swarms.	[9] Pham and Sholedolu, 2008
Simulated Annealing (SA) and Direct Annealing (DA)	Using SA or DA-based selection to update all elite and non-elite bees.	[10 Sadiq and Hamad, 2010; [11] Masmoudi *et al*., 2016
Genetic Algorithm (GA) and Differential Evolution (DE) Algorithm	Using mutation and crossover search operators and biased selection operator.	[12] Anh and Vi, 2023
Firefly Algorithm	Using firefly rules in global search.	[13] Gholami and Mohammadi, 2018
Crow Search Algorithm	Optimising in two stages.	[14] Deghbouch and Debbat, 2022
Fuzzy Logic	Transforming a multi-objective optimisation problem into a fuzzy inference system using fuzzy objective functions.	[15] Tolabi *et al.*, 2014
Artificial Neural Networks (ANN)	Using ANN to obtain a model and the BA to optimise its parameters.	[16] Ebrahimpoor *et al.*, 2019
Support Vector Machines (SVM)	Using the BA for feature selection before classification by SVM.	[17] Abdusalam *et al*., 2023
Convolutional Neural Networks (CNN)	Using the BA to optimise CNNs	[18] Alamri *et al*., 2022

V. THREE RECENT VERSIONS OF THE BEES ALGORITHM

This section highlights three new versions of the Bees Algorithm, of which one was designed to speed up its operation, another to simplify its parameter setting and the third, to achieve both simplicity and fast convergence. The first version is the dual population Bees Algorithm by Song et al [19], the second version is the two-parameter Bees Algorithm by Ismail et al [20] and the third version is the super-fast Bees Algorithm by Laili et al [21].

A. Dual-Population Bees Algorithm

Figure 3 [19] is a flowchart of this version of the Bees Algorithm. As its name implies, the algorithm uses two populations of bees, one with which to perform greedy (local) search and the other for increasing population diversity. The individual sizes of the two populations vary dynamically. The total size of the two populations is kept constant. The 'search' population contains bees that perform above the average of the two populations. The 'supplementary' population comprises below-average bees. Bees from the search population are demoted to the supplementary population if they fall below the average. Bees in the supplementary population can be probabilistically selected to create the population for the next iteration. Bees from the search population are demoted to the supplementary population if they fall below the average. Bees in the supplementary population can be probabilistically selected to create the population for the next iteration.

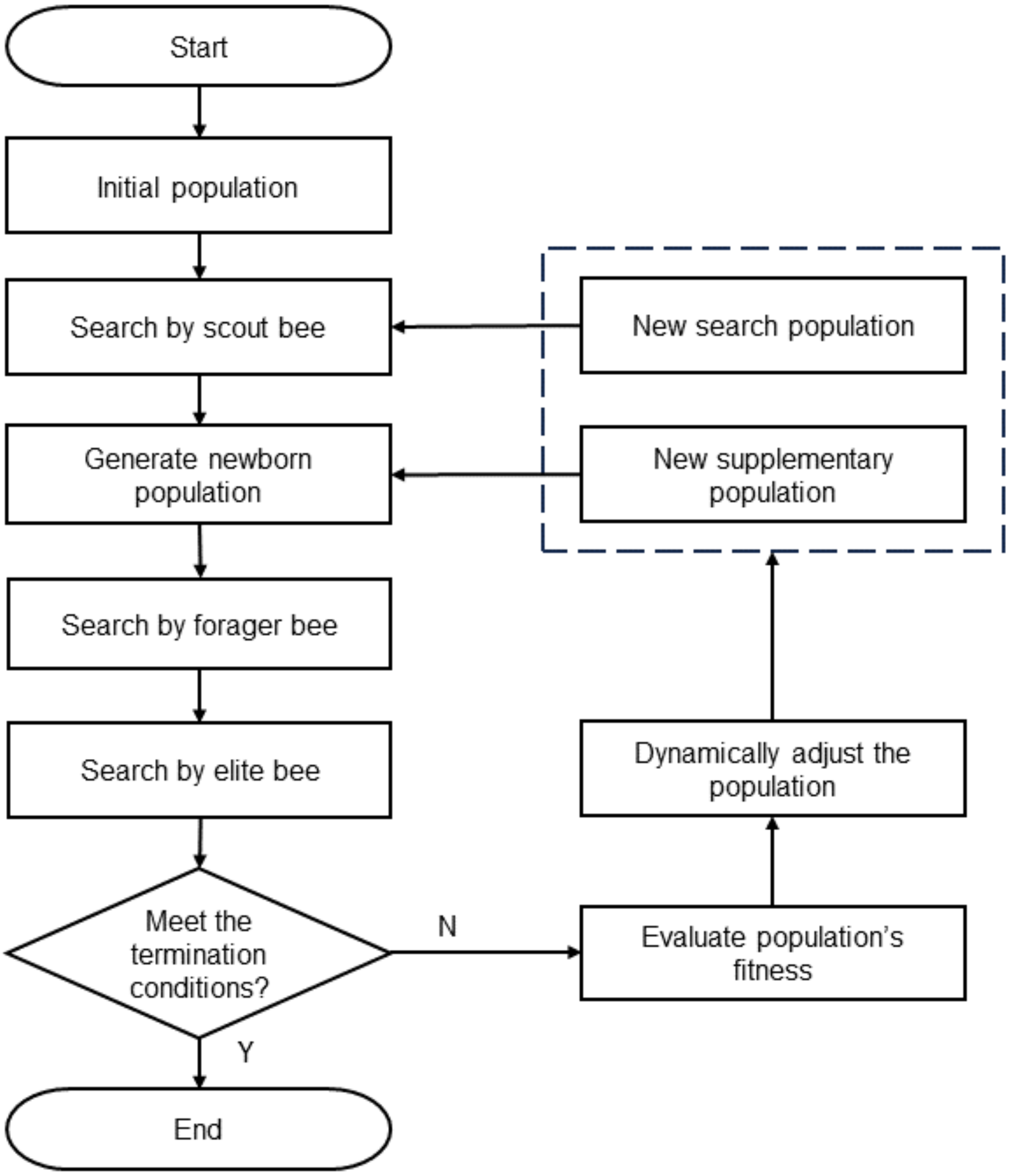
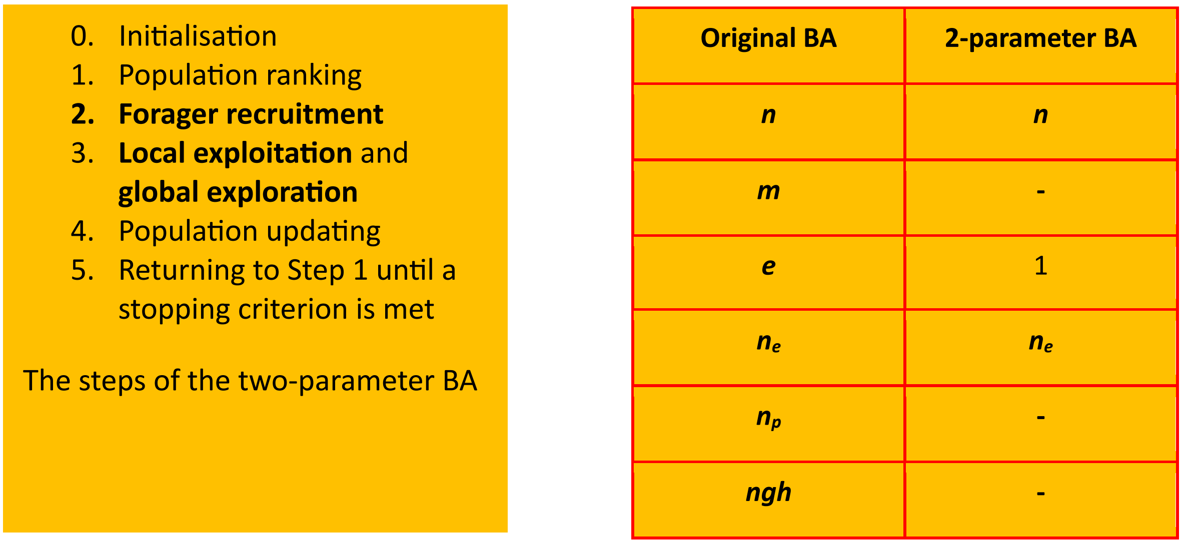


FIGURE 3. FLOWCHART OF DUAL-POPULATION BEES ALGORITHM [19]

It should be noted that this version of the Bees Algorithm is significantly different from the basic version described previously. This shows that Bees Algorithm research has virtually unlimited possibilities.

B. Two-parameter Bees Algorithm (BA2)

This algorithm follows the operation of the original Bees Algorithm but differs from it in two respects related to recruitment and exploitation. The algorithm is summarised in Figure 4 which also shows its parameters and those of the original algorithm [20].



0. Initialisation
1. Population ranking
2. **Forager recruitment**
3. **Local exploitation** and **global exploration**
4. Population updating
5. Returning to Step 1 until a stopping criterion is met

The steps of the two-parameter BA

Original BA	**2-parameter BA**
n	***n***
m	-
e	1
n_e	n_e
n_p	-
ngh	-

FIGURE 4. THE TWO-PARAMETER BA AND ITS USER-SELECTABLE PARAMETERS [20]

As with the original Bees Algorithm, the user specifies the number n of scout bees.

For the recruitment of foragers, the user sets the number ne of bees that follow the top scout bee (the single 'elite' bee that found the best flower patch or solution). A decreasing function is used to allocate foragers to follow the other scout bees. For example, the second best-performing bee could recruit half of ne, the third best-performing bee could have half of that number of foragers again, etc. Thus, there is no need for the user to choose m, e and np.

Another difference with the original Bees Algorithm is that in BA2, bees perform exploration and exploitation at the same time. Some followers focus on exploiting the area surrounding the solution discovered by the scout. Others wander around exploring other parts of the solution space. The higher the ranking of the scout (i.e., the better the flower patch that it found), the larger the number of followers that focus on exploiting the patch and the smaller the number of wanderers.

In BA2, the concept of neighbourhood is generalised to cover the whole solution space, and the user no longer has to choose ngh. For each scout, a probability function is used to determine the distribution of the foragers in the solution space, i.e., which followers focus on the discovered solution and which ones stray further away.

C. Super-fast Bees Algorithm

This is the simplest version of the Bees Algorithm developed to date. The algorithm uses only 3 bees. It conducts greedy local search with the best-performing bee and applies genetic operators (precedence preserving crossover and mutation) using the middle-ranking bee. The poorest-performing bee may be used for the crossover operation or may be discarded. If it is not crossed with the middle-ranking bee, the top-performing bee is used instead. In both cases, a new bee is randomly created to replace the poorest bee, thus maintaining population diversity.

Figure 5 [21] depicts the super-fast Bees Algorithm. The boxes labelled 1, 2 and 3 represent the actions of the best-performing, middle-ranking and worst bees, respectively. Local exploitation is conducted by the best bee, and global exploration, by the worst bee. The middle-ranking bee could be viewed as executing a hybrid of local and global search. As the population size is fixed and there is only one elite high-performing bee, the user does not have to specify n, m or e. No forager bees are employed for neighbourhood search and thus ne, np and ngh are not required either.

The super-fast Bees Algorithm, also called the Ternary Bees Algorithm because it uses three bees, is extremely efficient. This is a result of having a very small population. However, this could arguably be the algorithm's potential weakness causing it to converge quickly to false optima in some problems.

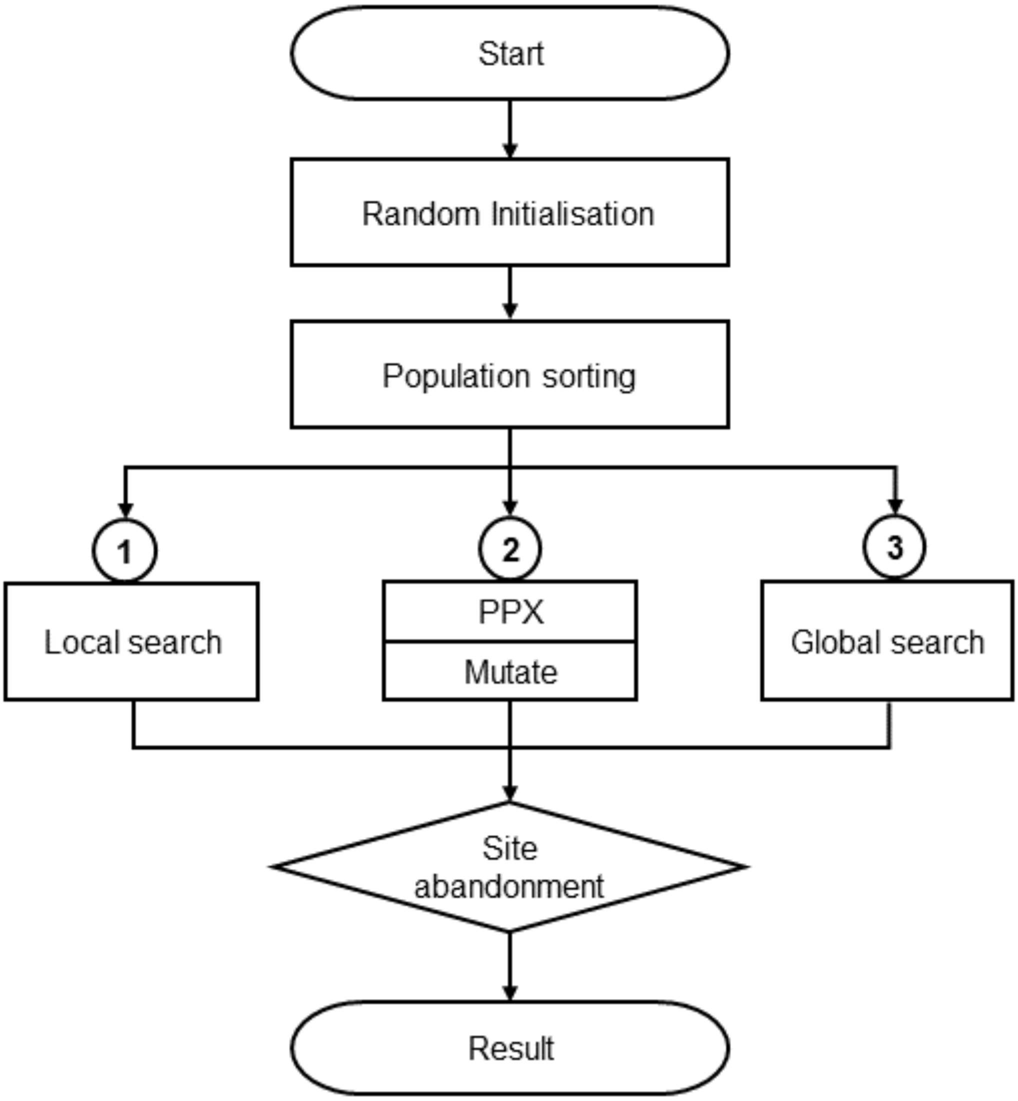


FIGURE 5. THE SUPER-FAST BEES ALGORITHM [21]

VI. SOME APPLICATIONS OF THE BEES ALGORITHM IN PRODUCTION AND MANUFACTURING

This section reviews some recent uses of the Bees Algorithm to solve optimisation problems in production and manufacturing. Both continuous and combinatorial optimisation problems have been addressed.

A. Continuous Optimisation

Packianather et al. [22] employed the Bees Algorithm to optimise process parameters for wire electrical discharge machining. The authors leveraged the algorithm to obtain the best combination of process parameters for the digital twin of the product being machined. Conte and d'Addona [23] applied the Bees Algorithm to tool wear identification and measurement during turning operations. The author's goal was to use the bees to define the contours of the wear area of a tool and locate the point of maximum wear.

Ay et al. [24] applied Bee-Miner, a cost-sensitive classification algorithm for data mining derived from the Bees Algorithm, to classify defects in manufacturing. Zeybek [25] trained a Long-Short-Term Memory (LSTM) deep learning network to predict the remaining useful life (RUL) of turbofan engines before they are available for remanufacturing with the help of a modified version of the super-fast Bees Algorithm.

A memory-based Bees Algorithm with a Lévy search facility was used by Shatnawi et al. [26] to perform multilevel thresholding of images, a basic operation required when processing images for automated visual inspection or general computer vision. Castellani et al. [27] have also adopted the Bees Algorithm in work relevant to vision systems. They used the Bees Algorithm to fit primitive shapes to point-cloud scenes for real-time 3D object recognition. Lan et al. [28] solved the problem of finding a spatial transformation that aligns two point clouds with an enhanced Bees Algorithm. They employed singular value decomposition to increase the search efficiency of the Bees Algorithm, achieving higher consistency, precision and robustness than the popular Iterative Closest Point method.

Öztürk et al. [29] used the Bees Algorithm to find the parameters (pulse frequency, amplitude and base) for a test to obtain the biaxial stress-strain curves required to control the pulsating hydroforming process to yield a uniform thickness distribution. Şahin and Çakıroğlu [30] employed a multi-objective Bees Algorithm to design PID control systems such as those for robotic equipment. They successfully found the PID parameters to minimise the settling time, rise time, overshoot, and system error all at once. Continuing with robotics applications, Kashkash et al. [31] modified the Bees Algorithm by adding a new initial population generation method and used the modified algorithm to find the shortest collision-free path for a mobile robot.

Mechanical design was performed by Acar et al. [32] using a version of the Bees Algorithm hybridised with Harmony Search. The hybrid algorithm which incorporated a harmony memory was employed to optimise the design of spherical four-link mechanisms for robot grippers.

B. Combinatorial Problems

Ang and Ng [33] optimised process plans for printed circuit board assembly. They combined the Bees Algorithm with TRIZ operators to minimise assembly time. Song et al. [19] used their dual-population Bees Algorithm to perform parallel machine scheduling with time windows. Production planning and scheduling for a sheet metal fabrication plant was accomplished by Wang et al. [34] using both the original Bees Algorithm and the version implementing the site abandonment strategy.

Li et al. [35] scheduled tasks and selected resources for a cloud-edge collaborative manufacturing environment using a dynamic Bees Algorithm. The operators in the algorithm were adjusted according to a set of indicators, and a parallel sorting scheme was adopted to accelerate the scheduling and selection of cloud-edge resources and collaboration modes.

Task allocation in a cloud remanufacturing environment was performed by Caterino et al. [36] with the help of a Bees Algorithm the parameters of which were selected following a full-factorial experiment. Their work confirmed the importance of increasing the number of scout bees and exploiting the best-performing and elite sites.

Remanufacturing applications of the Bees Algorithm also include the optimisation of product disassembly which is the first operation in any product remanufacturing process chain. For example, Liu et al. [37] employed a discrete Bees Algorithm simultaneously to optimise disassembly sequences and balance the disassembly line, with the analytic process network assigning weights to the different optimisation objectives. Hartono et al. [38] adopted a multi-objective Bees Algorithm to devise robotic disassembly sequence plans to achieve maximum profit while minimising energy consumption and greenhouse gas emissions.

The Bees Algorithm has been used for logistics and supply chain optimisation. Mastrocinque [39] implemented a tool based on the Bees Algorithm to optimise the supply chain for a bulldozer based on its Bill of Materials, finding that the Bees Algorithm performed better than Ant Colony Optimisation. Finally, Ismail and Pham [40] used the two-parameter Bees Algorithm to solve the capacitated vehicle routing problem. To speed up the solution, a decomposition approach was adopted whereby customers were first clustered before the optimal route was found for each cluster.

VII. CONCLUSION

This paper has presented the main features of the original Bees Algorithm and some enhancements to it. Three new versions have also been reviewed. The paper concludes by giving examples of the use of the Bees Algorithm in production and manufacturing. There are many more potential applications, especially with the advent of the recently developed simpler and more efficient versions.

It has been seen how the original Bees Algorithm mimics the foraging behaviour of bees in nature where three types of bees exist. Scout bees perform exploration then go back to the hive to recruit forager bees to follow them and help exploit the discovered food sources. Those foragers are initially observer bees in the hive. They spring into action after watching the waggle dance by the scout bees.

It is hoped that, like observer bees, some readers of the paper will spring into action and become forager bees. In the field of metaheuristic optimisation, there is so much more to explore and exploit. The new versions outlined in this paper have only been used for a limited number of problems in production engineering. They could readily be applied to many more problems.

For researchers who have used other optimisation algorithms, comparing their results with those obtained using the Bees Algorithm will confirm the effectiveness of this powerful metaheuristic technique. Further research avenues also beckon, such as hybridising the Bees Algorithm with other tools and performing deep theoretical analysis of the resulting hybrid algorithms to understand their true areas of strength.

1. D T Pham, A Ghanbarzadeh, E Koc, S Otri, S Rahim and M Zaidi (2005) The Bees Algorithm, Technical Note, Manufacturing Engineering Centre, Cardiff University, UK

2. D T Pham, A Ghanbarzadeh, E Koç, S Otri, S Rahim and M Zaidi (2006). The Bees Algorithm - a novel tool for complex optimisation problems, Proceedings of the 2nd Virtual International Conference on Intelligent Production Machines and Systems, D T Pham, E E Eldukhri and A J Soroka (eds), Elsevier (Oxford) (2006), 454-460, ISBN 0-08-045157-8

3. Q T Pham, D T Pham and M Castellani (2012). A modified bees algorithm and a statistics-based method for tuning its parameters, Proceedings of Institution of Mechanical Engineers, Part I, Volume 226, 287-301, ISSN 0954-4054

4. M C Ang, D T Pham and K W Ng (2009). Application of the Bees Algorithm with TRIZ-inspired operators for PCB assembly planning, Proceedings of the 5th Virtual International Conference on Innovative Production Machines and Systems, D T Pham, E E Eldukhri and A J Soroka (eds), Whittles (Dunbeath, Caithness) and CRC Press (Boca Raton, FL), 405-410, ISBN 978-1-84995-007-7W

5. D T Pham and A Haj Darwish (2010). Using the Bees Algorithm with Kalman filtering to train an Artificial Neural Network for Pattern Classification, Proceedings of Institution of Mechanical Engineers, Part I, Volume 224, 885-892, ISSN 0959 6518

6. A Hussein, S Sahran and S N H Sheikh Abdullah (2017). The variants of the Bees Algorithm (BA): a survey. Artif Intell Rev, 47, 67–121.

7. A H Ismail (2021). Enhancing the Bees Algorithm using the Traplining Metaphor, PhD Thesis, University of Birmingham, UK.

8. M Packianather, M Landy and D T Pham (2009). Enhancing the speed of the Bees Algorithm using pheromone-based recruitment classifiers, Proc IEEE 7th International Conference on Industrial Informatics (INDIN 2009), Cardiff, UK, 789-794, ISBN: 978-142443760-3

9. D T Pham and M Sholedolu (2008). Using a hybrid PSO-Bees Algorithm to train neural networks for wood defect classification, Proceedings of the 4th Virtual International Conference on Innovative Production Machines and Systems, D T Pham, E E Eldukhri and A J Soroka (eds), Whittles (Dunbeath, Caithness) and CRC Press (Boca Raton, FL), 385-390, ISBN 978-1-4398-0117-8

10. A T Sadiq and A G Hamad (2010). BSA: A hybrid bees' simulated annealing algorithm to solve optimization and NP-complete problems. Eng. & Tech. Journal, 28(2), 271-281

11. M A Masmoudi, M Hosny, K Braekers and A Dammak (2016). Three effective metaheuristics to solve the multi-depot multi-trip heterogeneous dial-a-ride problem, Transportation Research Part E: Logistics and Transportation Review, 96, 60-80, ISSN 1366-5545, https://doi.org/10.1016/j.tre.2016.10.002

12. T T M Anh and T D Vi (2023). Hybrid Genetic-Bees Algorithm in Multi-layer Perceptron Optimization. In: Saraswat, M., Chowdhury, C., Kumar Mandal, C., Gandomi, A.H. (eds) Proceedings of International Conference on Data Science and Applications. Lecture Notes in Networks and Systems, 552, Springer, Singapore. https://doi.org/10.1007/978-981-19-6634-7_11

13. J Gholami and S Mohammadi (2018). A Novel Combination of Bees and Firefly Algorithm to Optimize Continuous Problems, Proceedings of 8th International Conference on Computer and Knowledge Engineering (ICCKE), Mashhad, Iran, 40-46, https://doi.org/10.1109/ICCKE.2018.8566263

14. H Deghbouch and F Debbat (2022). Hybrid algorithm for node deployment with the guarantee of connectivity in wireless sensor networks, Informatica, 46(8), 147-164, https://doi.org/10.31449/inf.v46i8.3370

15. H B Tolabi, M H Ali, S B M Ayob and M Rizwan (2014). Novel hybrid fuzzy-Bees algorithm for optimal feeder multi-objective reconfiguration by considering multiple-distributed generation, Energy, 71, 507-515, https://doi.org/10.1016/j.energy.2014.04.099

16. S Ebrahimpoor, V Kiarostami, M Khosravi, M Davallo and A Ghaedi (2019). Bees metaheuristic algorithm with the aid of artificial neural networks for optimization of acid red 27 dye adsorption onto novel polypyrrole/SrFe12O19/graphene oxide nanocomposite. Polym. Bull., 76, 6529–6553, https://doi.org/10.1007/s00289-019-02700-7

17. O Abdusalam, F Anayi and M Packianather (2023). Three-Phase Power Transformer Fault Diagnosis Based on Support Vector Machines and Bees Algorithm, Proceedings 2023 IEEE 3rd International Conference in Power Engineering Applications (ICPEA), Putrajaya, Malaysia, 145-150, https://doi.org/10.1109/ICPEA56918.2023.10093147

18. N M H Alamri, M Packianather and S Bigot (2022). Deep Learning: Parameter Optimization Using Proposed Novel Hybrid Bees Bayesian Convolutional Neural Network, Applied Artificial Intelligence, 36(1), https://doi.org/10.1080/08839514.2022.2031815

19. Y Song, L Xing and Y Chen (2023). Application of the dual-population Bees Algorithm in a parallel machine scheduling problem with a time window. In: Pham, D.T., Hartono, N. (eds) Intelligent Production and Manufacturing Optimisation—The Bees Algorithm Approach. Springer Series in Advanced Manufacturing. Springer, Cham. https://doi.org/10.1007/978-3-031-14537-7_14

20. A H Ismail, W Ruslan and D T Pham (2023). A user-friendly Bees Algorithm for continuous and combinatorial optimisation, Cogent Engineering, 10:2, https://doi.org/10.1080/23311916.2023.2278257

21. Y Laili, F Tao, D T Pham, Y Wang and L Zhang (2019). Robotic disassembly re-planning using a two-pointer detection strategy and a super-fast bees algorithm, Robotics and Computer-Integrated Manufacturing, 59, 130-142, https://doi.org/10.1016/j.rcim.2019.04.00

22. M Packianather, T Alexopoulos and S Squire, S. (2023). The application of the Bees Algorithm in a digital twin for optimising the wire electrical discharge machining (WEDM) process parameters. In: Pham, D.T., Hartono, N. (eds) Intelligent Production and Manufacturing Optimisation—The Bees Algorithm Approach. Springer Series in Advanced Manufacturing. Springer, Cham. https://doi.org/10.1007/978-3-031-14537-7_3

23. S Conte and D M d'Addona (2023). Bees Algorithm models for the identification and measurement of tool wear. In: Pham, D.T., Hartono, N. (eds) Intelligent Production and Manufacturing Optimisation—The Bees Algorithm Approach. Springer Series in Advanced Manufacturing. Springer, Cham. https://doi.org/10.1007/978-3-031-14537-7_7

24. M Ay, A Baykasoglu, L Ozbakir and S Kulluk (2023). A case study with the BEE-Miner Algorithm: defects on the production line. In: Pham, D.T., Hartono, N. (eds) Intelligent Production and Manufacturing Optimisation—The Bees Algorithm Approach. Springer Series in Advanced Manufacturing. Springer, Cham. https://doi.org/10.1007/978-3-031-14537-7_4

25. S Zeybek (2023). Prediction of the remaining useful life of engines for remanufacturing using a semi-supervised deep learning model trained by the Bees Algorithm. In: Pham, D.T., Hartono, N. (eds) Intelligent Production and Manufacturing Optimisation—The Bees Algorithm Approach. Springer Series in Advanced Manufacturing. Springer, Cham. https://doi.org/10.1007/978-3-031-14537-7_21

26. N Shatnawi, S Sahran and M F Nasrudin (2023). Memory-based Bees Algorithm with Lévy Flights for multilevel image thresholding. In: Pham, D.T., Hartono, N. (eds) Intelligent Production and Manufacturing Optimisation—The Bees Algorithm Approach. Springer Series in Advanced Manufacturing. Springer, Cham. https://doi.org/10.1007/978-3-031-14537-7_11

27. M Castellani, L Baronti, S Zheng and F Lan (2023). Shape recognition for industrial robot manipulation with the Bees Algorithm. In: Pham, D.T., Hartono, N. (eds) Intelligent Production and Manufacturing Optimisation—The Bees Algorithm Approach. Springer Series in Advanced Manufacturing. Springer, Cham. https://doi.org/10.1007/978-3-031-14537-7_6

28. F Lan, M Castellani, Y Wang and S Zheng (2023). Global optimisation for point cloud registration with the Bees Algorithm. In: Pham, D.T., Hartono, N. (eds) Intelligent Production and Manufacturing Optimisation—The Bees Algorithm Approach. Springer Series in Advanced Manufacturing. Springer, Cham. https://doi.org/10.1007/978-3-031-14537-7_8

29. O Öztürk, M A Şen, M Kalyoncu and H S Halkacı (2023). An application of the Bees Algorithm to pulsating hydroforming. In: Pham, D.T., Hartono, N. (eds) Intelligent Production and Manufacturing Optimisation—The Bees Algorithm Approach. Springer Series in Advanced Manufacturing. Springer, Cham. https://doi.org/10.1007/978-3-031-14537-7_5

30. M Şahin and S Çakıroğlu (2023). Automatic PID tuning toolkit using the Multi-Objective Bees Algorithm. In: Pham, D.T., Hartono, N. (eds) Intelligent Production and Manufacturing Optimisation—The Bees Algorithm Approach. Springer Series in Advanced Manufacturing. Springer, Cham. https://doi.org/10.1007/978-3-031-14537-7_9

31. M Kashkash, A Haj Darwish and A Joukhadar (2023) . A new method to generate the initial population of the Bees Algorithm for robot path planning in a static environment. In: Pham, D.T., Hartono, N. (eds) Intelligent Production and Manufacturing Optimisation—The Bees Algorithm Approach. Springer Series in Advanced Manufacturing. Springer, Cham. https://doi.org/10.1007/978-3-031-14537-7_12

32. O Acar, H Sağlam and Z Şaka (2023). The effect of Harmony Memory integration into the Bees Algorithm. In: Pham, D.T., Hartono, N. (eds) Intelligent Production and Manufacturing Optimisation—The Bees Algorithm Approach. Springer Series in Advanced Manufacturing. Springer, Cham. https://doi.org/10.1007/978-3-031-14537-7_10

33. M C Ang and K W Ng (2023). Minimising printed circuit board assembly time using the Bees Algorithm with TRIZ-inspired operators. In: Pham, D.T., Hartono, N. (eds) Intelligent Production and Manufacturing Optimisation—The Bees Algorithm Approach. Springer Series in Advanced Manufacturing. Springer, Cham. https://doi.org/10.1007/978-3-031-14537-7_2

34. C Wang, T Chen and Z Li (2023). Method for the production planning and scheduling of a flexible manufacturing plant based on the Bees Algorithm. In: Pham, D.T., Hartono, N. (eds) Intelligent Production and Manufacturing Optimisation—The Bees Algorithm Approach. Springer Series in Advanced Manufacturing. Springer, Cham. https://doi.org/10.1007/978-3-031-14537-7_13

35. Y Li, C Peng, Y Laili and L Zhang (2023). A parallel multi-indicator-assisted dynamic Bees Algorithm for cloud-edge collaborative manufacturing task scheduling. In: Pham, D.T., Hartono, N. (eds) Intelligent Production and Manufacturing Optimisation—The Bees Algorithm Approach. Springer Series in Advanced Manufacturing. Springer, Cham. https://doi.org/10.1007/978-3-031-14537-7_15

36. M Caterino, M Fera, R Macchiaroli and D T Pham (2023). Task optimisation for a modern cloud remanufacturing system using the Bees Algorithm. In: Pham, D.T., Hartono, N. (eds) Intelligent Production and Manufacturing Optimisation—The Bees Algorithm Approach. Springer Series in Advanced Manufacturing. Springer, Cham. https://doi.org/10.1007/978-3-031-14537-7_20

37. J Liu, Q Liu, Z Zhou, D T Pham, W Xu and Y Fang (2023). Collaborative optimisation of robotic disassembly planning problems using the Bees Algorithm. In: Pham, D.T., Hartono, N. (eds) Intelligent Production and Manufacturing Optimisation—The Bees Algorithm Approach. Springer Series in Advanced Manufacturing. Springer, Cham. https://doi.org/10.1007/978-3-031-14537-7_18

38. N Hartono, F Javier Ramírez and D T Pham (2023). Optimisation of robotic disassembly sequence plans for sustainability using the multi-objective Bees Algorithm. In: Pham, D.T., Hartono, N. (eds) Intelligent Production and Manufacturing Optimisation—The Bees Algorithm Approach. Springer Series in Advanced Manufacturing. Springer, Cham. https://doi.org/10.1007/978-3-031-14537-7_19

39. E Mastrocinque (2023). Supply chain design and multi-objective optimisation with the Bees Algorithm. In: Pham, D.T., Hartono, N. (eds) Intelligent Production and Manufacturing Optimisation—The Bees Algorithm Approach. Springer Series in Advanced Manufacturing. Springer, Cham. https://doi.org/10.1007/978-3-031-14537-7_17

40. A H Ismail and D T Pham (2023). Bees traplining metaphors for the Vehicle Routing Problem using a decomposition approach. In: Pham, D.T., Hartono, N. (eds) Intelligent Production and Manufacturing Optimisation—The Bees Algorithm Approach. Springer Series in Advanced Manufacturing. Springer, Cham. https://doi.org/10.1007/978-3-031-14537-7_16

International Journal of Advances in Production Research

Industrial Augmented Reality: A Framework for Defining Requirements

Jan Luca Siewert[a,1], Detlef Gerhard[a,2*]

[a] *Digital Engineering Chair, Ruhr University Bochum, Germany*

[1]`Jan.Siewert@ruhr-uni-bochum.de` [2]`Detlef.Gerhard@ruhr-uni-bochum.de`

Abstract— **Augmented Reality (AR) boasts a wide array of applications throughout the entire product lifecycle; however, its adoption in industrial settings is often impeded by factors such as high setup costs, poor system integration, and limited modality. Despite these challenges, current Industrial Augmented Reality (IAR) applications exhibit a significant overlap in components used for information gathering and visualization. This paper describes the current state of modern IAR architectures and introduces a novel framework for the definition of requirements specific to IAR ap-plications. This draws upon the principles of human-cantered design to provide a structured approach for integrating IAR more effectively in industrial contexts. Finally, a case-study shows how the framework can be used to help imple-ment an assistance system for assembly of a simple product.**

Keywords— **Augmented Reality, Industrial Augmented Reality, Requirements Engineering, Human Machine Interaction, System Design**

I. INTRODUCTION

Industrial Augmented Reality (IAR) has many applications across the whole product lifecycle, including as-sembly support, service and maintenance, or marketing applications [1]. Various studies show that the use of IAR can be advantageous, for example in reducing training times or the number errors when performing complex tasks [2], [3], [4]. Despite more capable AR hardware [5] and recent progress in presentation and tracking quality [6], industrial adaption for IAR remains low [7]. Future research should therefore focus more on organizational issues instead of technology alone, to make implementing AR more cost-effective and reducing training times [7].

The biggest issues relate to available software as well as required implementation and modification time to integrate them into existing processes. This is especially true for small and medium-sized companies (SMEs). One strategy might be introducing more modular solution that reduce these required implementation times and could help boaster adoption for SMEs in particular [5].

Even when working with those modular architectures, the afore mentioned limitations and trade-offs makes the requirement definition for these applications a special challenge in and itself. In a modular system, one has to decide if existing modules can be reused or new ones have to be developed, and if so, how functionality should be bundled into independent modules. This contribution therefore presents a methodology for implementing a mod-ular IAR system out of reusable components. To describe potential applications for such systems, IAR use-cases are described and classified in more detail. Then, potential approaches and architectures for modular IAR systems are described. Based on an overview of the current literature regarding requirement definition for AR and IAR systems, the mentioned methodology is developed. This includes both a proposed process based on DIN EN ISO 9241-210 [8] and some considerations and essential inquiries one should ask when defining requirements as a starting point. Finally, the proposed methodology is applied to a case study.

II. STATE OF THE ART AND PRIOR RESEARCH

To be able to describe a framework for requirement definition, first, the current state of the art and research gaps are described. To be able to generalize and break down monolithic IAR applications into reusable components, a common terminology to describe the usage context has to be established. For that, classification schemes for IAR use-cases are presented. Then, different approaches for modular architectures are presented. Finally, com-mon requirements and frameworks for requirement engineering for IAR are described.

A. State of the Art

Today's AR hardware is already adequate to support even large-scale AR applications [5]. A lot of progress also has been made in presentation quality, as well as tracking and registration [6]. Context-aware AR applications have been proposed numerous times, for example with the introduction of deep-learning methods. The process made in display technologies, presentation quality, as well as tracking and registration is highlighted in [6]. While content can be presented using 3D models, texts, and symbols, the most appropriate representation used is highly dependent on the operator's preference [6], [9], [10].

*Corresponding Author

Today, there are four predominant presentation technologies used in AR: video-see-through (VST) headsets, where the operator sees both a camera feed and virtual information through displays; optical see through (OST), where the operators directly sees the environment; mobile AR, where smart devices are used to render the camera feed and overlays; and projection based AR, where information is projected directly in top of the environment. All of these have different advantages and disadvantages, as [11] show with expert interviews. VST allows hands-free operation and high-quality overlays, while raising ergonomic problems. Furthermore, because the environment is only observed through screens and cameras, it raises both safety and security issues. On the other hand, OST allows direct perception of the environment, with the downside of decreased quality of virtual overlays. Ergonomic issues prevalent in these kinds of head mounted devices (HMDs), as well. Mobile AR allows known interactions. However, hands-free operation is an issue, as the operator constantly has to place and pick up the device. Projection based AR offers good compromises, allows hands-free operation, undisrupted overview of the environment, and no additional weight or devices to wear. Issues are its inflexibility and possible occlusions [11].

The different trade-offs can be seen on the example of maintenance specifically. There, mobile AR has a lot of advantages, like the ease-of-use and reliability. HMDs are therefore most useful when hands-free activity is required. That is usually during after an initial information phase during the actual execution [9].

While a lot of research has been done to improve model recognition and tracking, even recent AR applications mostly rely on a marker-based approach because of better reliability. Besides this and the afore mentioned ergonomic issues, organizational challenges, like data integration and content authoring are major challenges for the adoption of AR in the industry [12].

When the accuracy requirements are higher than today’s tracking and registration methods allow, combining external sensors becomes a viable alternative [9]. This is also often used in industrial applications [13]. Industrial experiences also highlight that HMDs are mostly not suitable for industrial use while that mobile and projection-based AR coupled with external tracking becomes a viable solution [13]. Other external sensors that might be coupled with IAR systems might be used for data acquisition, to monitor the environment, but also depth cameras for motion capture [9]. Additionally, industrial experiences shows that workers prefer and require different level of support or else may feel hindered by an assistant system [13], [9].

B. Industrial AR Use-Cases

Röltgen and Dumitrescu present a systematic literature review about the different use-cases of IAR. They define a total of 26, like marketing, product design, assembly, or maintenance support. They also present a classification scheme that can be used to describe a use-case. To describe the context for the system, they define four types of actions a system can support: inform, execute, plan, and control. It is influenced by the field of application in the product lifecycle the system is used. For the lifecycle phase they propose to use (1) procurement, (2) engineering, (3) production, (4) logistics, (5) maintenance, (6) decommissioning, and (7) training. Additionally, they differentiate whether the system effects the virtual or real world. Technological factors of the systems focus on the aim of the augmentation (performance enhancement, qualification, or enhanced perception). This influences the spatial location of information (user, object, or environment), the temporal context of the augmentation (whether it refers to information from the past, present, future, or is fiction). Finally, they differentiate if it is desirable to separate virtual information from reality or not [14].This classification scheme spans eight dimensions with a lot of possible options, which is in contrast to the 16 observed applications. Furthermore, the classification introduces both the context and the technological implementation, which makes it suitable to describe existing applications, but is not as useful when trying to describe to goal of a future application, alone.

Therefore, we have proposed an alternative classification based on this [15]. Focus is the description of the use-cases alone. This classification scheme keeps both the supported action and the supported lifecycle phase. These are usually well-defined when implementing an assistant system. As third and final dimension, the authors propose to use the desired level of support. As discussed above, using a suitable level of support for the specific workers is an important factor to consider. The degree of support might range from low, which includes visualizing information while the operator is responsible for decision-making, to high where specific actions to perform are recommended by the system. A special case in this category is collaboration with a remote expert. This is considered part of this category because the support is not offered by the system, which is only responsible for facilitating communication, but by using the expert’s expertise.

C. Modular IAR Architectures

Modular architectures for AR applications have various advantages for the different roles involved. As the implementation work can be broken down into the different domains, specialists can effectively work on them individually. Functionality can be evaluated more effectively and generalized into reusable components. This allows for quicker prototypes and testing, as well as for increased flexibility to address the needs of specific users [16]. The contribution proposes to create reusable components along the traditional three-layer architecture of software systems. The view layer is separated into interaction modules, presentation modules, and tracking modules that incorporate their data into a universal world module, that controls placement of virtual information. This approach allows the combination of multiple sources for tracking information, like AR tracking and GPS. The application layer is responsible for integrating the data, for example by providing the necessary algorithms for sensor fusion. A task flow engine models a finite-state-machines, where each state is associated with necessary support documents that are shown to the user while the state is active. Context aware services abstract external services that might be added to the AR system, like printing. The user interface is modelled on top of a specific markup language and is also associated with a state machine, comparable to the task-flow engine [16].

For mobile devices specifically, the three core components are responsible for tracking, rendering a camera image, and rendering 3D components [17]. In their work for mobile devices, they decouple these functionalities to allow an easier implementation and adoption for content. Both works show the advantages such a modular architecture can have for adaptability and flexibility. However, the fundamental ways AR systems are developed has shifted in recent years. The described capabilities are provided by software development kits (SDKs). This includes both mobile devices, for which various available SDKs bundle tracking and rendering functionality, but even more so for HMDs, where this core functionality is provided directly by the operating systems.

Recently proposed architectures focus on IAR applications and shifted away from the technical implementation to data integration and communication between different applications. Instead of dividing the technical implementation on a device into the three layers, one can separate between the visualization and interaction layer, a data transport layer, and a data acquisition layer as a viable alternative for IAR applications [18]. The data acquisition layer could connect to existing PLM, ERP, and IoT systems, while the data transport layer could exist as edge computing and be responsible for gateways and data proxies to Cache dynamically created IAR information. Finally, the data visualization layer is responsible for interactions with the operator [18].

To facilitate communication between different services in IAR applications, a distributed and service-oriented architecture can be used [19]. Different services, like object recognition, barcode decoding, or knowledge management, register availability on a service registry. Data from these services can be accessed through a companion device, that then passes the data on to the actual AR device. The architecture focuses on service and maintenance tasks across multiple companies. The work plan is described using Business Process Model and Notation (BPMN) [18]. The proposed architecture shows how modularity can facilitate shared implementation work across multiple companies, but focuses on data flows and communication, not on the technical implementation of the AR content.

Content for AR applications can be shown in different ways. The different presentation devices, VST, OST, mobile AR, and projection, all have different advantages and limits, so that the most suitable devices for a given use case must choose on a case-by-case basis [11]. Studies also show performance differences between content representations. For example, presenting 3D models are especially useful when highlighting blind spots where the operator's view is obstructed. It decreased the completion time compared to 2D renderings and other content representation times [20].

A comprehensive literature review for content representation describes IAR content as a combination of a feature, the asset or content representation used, and the anchor that describes its position in space [21]. The content types that can be used can be text, signs, photography, video, drawings, technical drawing, a product model, or auxiliary models, like arrows [21]. Similar to the output device, the most appropriate representation to used depends on many factors. The most common ones are text, symbols as a combination of 2D signs and 3D auxiliary models, and product models. The fact that the same data can be represented differently allows for more flexibility in the development of AR systems. As not all output devices support all data representations, e.g. a projector not supporting full 3D models, alternatives may be providing a 2D drawing or rendering instead.

Based on this works and assumptions, we have developed a dynamic architecture based on reusable components [22]. Based on the three layers established in [16] and [18], the architectures devices a data, application, and view layer. The data layer is responsible for data persistency and interfaces to other enterprise systems. It may implement a proxy to cache dynamic content and provide services that may be used to convert data representations by changing data formats, or by transforming one representation to another, e.g. by rendering a 3D model into an image. These services registry their availability into a service registry, as established in [19]. The application layer is responsible for executing the work plan. It is modelled as a state-machine. Each state describes the desired AR annotation that should be shown to the operator, while state transitions are triggered by occurring events. Furthermore, the application checks if available components are in fact capable of executing the work plan with its required data representations, and schedules data conversions on the data layer when necessary. The view layer is divided into components responsible for interaction and presentation. While interaction systems emit events, e.g. by user interaction or through automated systems or components, like buttons, voice interactions, or external sensors, presentation components show AR annotations to the operator. An AR annotation combines an anchor point with multiple data representations. When an annotation contains more than one representation, the most suitable one is used in the current environment. The view layer communicates with the application layer through MQTT. Presentation components subscribe to a combination of anchor point and used data representation. By utilizing MQTT topics, data is only forwarded to components that support it.

This architecture differs from other proposed solution in two ways:

1. it explicitly supports not only full AR on HMDs or mobile devices, but spans from these AR implementations to already established pick-to-light systems. Because the required data representations and anchor points can be fulfilled through multiple components, a pick-to-light system may provide the object anchor, while 2D data representations are presented on a monitor.
2. the proposed components make use of anchor points and data representation for typical IAR applications, while leaving the technical implementation of those as a black-box. This middle-ground maintains a high flexibility in the developed systems, while establishing a common ground for defining and describing the individual components.

To describe data representation, the architecture focuses on the technical aspects. In contrast to [21], photography, drawings, and technical drawings are combined into an image representation type while both 2D signs and 3D auxiliary models are combined into a symbol representation. The main motivation behind the change is that the representation type is used to define capabilities of components instead of the effect it has on the worker. As shown in figure 1, the anchor points used are adapted to IAR applications as well. Detached content, or content without an anchor, is shown in screen-space, for example in a head-up-display style in an HMD or at another, ease-to-read position chosen by the component. Content can be anchored spatially, usually at a specific product, or by anchoring it indirectly on an object using a given offset. Content may also only reference a given location. Additionally, content can be anchored to either a specific instance of an object or a given object type. The latter is especially important when indicating the location of a storage container.

Before implementing an IAR system, the desired functionalities and requirements have to be defined. To support this process, various frameworks exist for systems in human-machine-interactions in general and IAR in particular. These are presented in the next section.

D. Requirement Definition for Industrial AR

When designing IAR systems, various factors need to be considered to ensure the project's success. Aspects include technical considerations for the output hardware, how information is presented, and which information is required, how the operator can interact with the system, authoring and availability of content and data integration and processing. For that, a human-centred design should be followed [23]. Additional, common requirements for IAR applications include cost efficiency, data security, established regulations and laws, including ergonomics. For the first aspect, short setup times, the overall reliability of the system, as well as the accuracy of shown information, which are often gathered in near-real-time, are crucial [24].

Based on expert workshops, some general requirements for a maintenance support application are described in [25]. First, one of the biggest advantages of the technology is portable access to relevant information. This includes information about required tools, materials, and spare parts. The Spatial information in IAR system can for example be used to assist orientation in an assembly. Workflow guidance can be given more effectively using 3D animation, while the hands-free nature of most HMDs makes taking notes and pictures easier. Access to live telemetry data or cross-referencing existing cases are other advantages of permanent access to information. Finally, video calls with experts allow more effective work procedures. Other often requested features are an offline mode or recording of statistical data [21].

[26] describes factors that affect industrial adaption in four categories: task, workforce, context, and technology. Considerations regarding the tasks, it should be sufficiently complex to justify the overhead of an AR solution. Tasks that benefit from remote work and off-site experts benefit greatly from AR. Information that is presented by an IAR system needs to be established and codified and required information for defining and describing the tasks needs to be available in the first place. Additionally, the skills of the workforce need to be considered. Providing simple instructions to expert technicians does not benefit them, neither does an instruction that requires some expertise for novice workers. The right balance between technical and practical abilities and shown information needs to be found. Additionally, digital skills and the technology acceptance level of the workforce needs to be taken into account. For the work context, leadership and organizational processes to be respected so that the solution does not only integrate into established software systems but processes as well. Additional considerations are accessibility, connectivity, comfort, but also tool availability and ease of use as technological success factors. [26]

When data and instructions are only available on paper, [27] proposed a simple process for converting those into interactive AR work plans. First, the existing manual is analysed. Then, the work is divided in atomic actions. These can either be concepts or references, that should describe information using simple imagery and text, or actionable tasks. These should use annotated images, descriptive graphical symbols, and descriptions of simple texts. Finally, the actions should be grouped and organized with a single message per entity, consistent texts, and recurring symbols [27].

Based on the general requirements and [8], a process for defining requirements for IAR systems is developed in [28]. First, the user context and task in analysed and the desired hardware is selected. Then, user requirements are collected based upon first tests with the hardware, and suitable interaction schemes are selected. The designed solution is developed and evaluated. For evaluation, two tests in laboratory environments and one in the field of application are proposed [28].

For evaluation, [29] presents an overview of different usability studies for AR in general and IAR in particular. For those, most studies are recruiting young university studies and are mostly lab-based tests. The NASA TLX score is the de-facto standard for testing usability, while time and error rate and accuracy might be evaluated as secondary metrics [29].

As in [28], [30] puts a lot of focus on the hardware-selection and developed criteria for evaluation HMDs in particular. Relevant factors are the cost, weight, and technical aspects, like the field-of-view battery power, camera resolutions, as well processor speed, available storage, and memory. Other important factors are available programming interfaces, availability, and connectivity to external sensors or audio devices, as well as the used OS. Finally, ratings for dust and water resistance might be critical depending on the use-case [30].

The most comprehensive framework for requirement definition for IAR system is presented based on a literature review and expert interviews in [31]. First, they define 21 task keywords for atomic actions in maintenance and assembly task, for example remove, push, or measure, in the categories locate, check, operate, or other tasks. The main information type that can be indicated to the user might me operation, indication of a goal, or showing a sample. This is used to define a usage context.
They define the types for coordinate systems or anchor points for virtual information: the coordinate system of the HMD in the form of a head-up-display; the world; a body part; an object; or a combination of the HMD and the object or of two objects. For data representations, they follow [21]. Then, they outline 18 conditions that can be used to select an appropriate combination of data representation and anchor point based on the usage context. Some examples are to use only texts and images for simple instructions, to not use an object anchor when the system requires accuracy below 1 cm, or to not show product models that do not fit into the field-of-view of the headsets.
Main limitations of the study are its focus on HMDs, the fact that preferences and experiences of the user are not taken into account, and that it only describes requirements for the overall system.

To build on the advantages for modular architectures outlined in the previous works, for example by using the architecture described in [22], the requirements must not only be defined for the overall system, but there needs to be a process for decomposing the system and defining the individual components.

III. PROPOSED FRAMEWORK

As described in the previous section, splitting monolithic IAR systems into reusable components has several advantages. In previous works, the authors have shown the advantages of a three-layer architecture with separate interaction and presentation modules in the view layer [22]. The interaction events are responsible for triggering events that advance a work plan, while presentation modules show annotations to the worker. Annotations are a combination of a data representation, like images, videos, or 3D models, and anchor points. An annotation can be shown in a single representation component or split between them. The anchor point, e.g. a specific object or other position in space, is the spatial information attached to the information.

When designing and implementing such a modular system, a developer must select the required events, anchor points, and suitable representations. As described earlier, there is no single best system for a given use-case. Rather, the solution depends on available data, expertise, existing systems and experiences, and the preferences and requirements the worker that uses the IAR system. The presented methodology for defining and implementing an IAR system out of reusable components is based upon best practices and experiences from the literature, adapted to the component-based architecture. It follows the standard for human centred design for interactive systems [8] to describe the context and requirements for the overall system, define required events, anchor points, and data representation the system should be capable of.

The basic process and its relation to [8] is presented in figure 2.

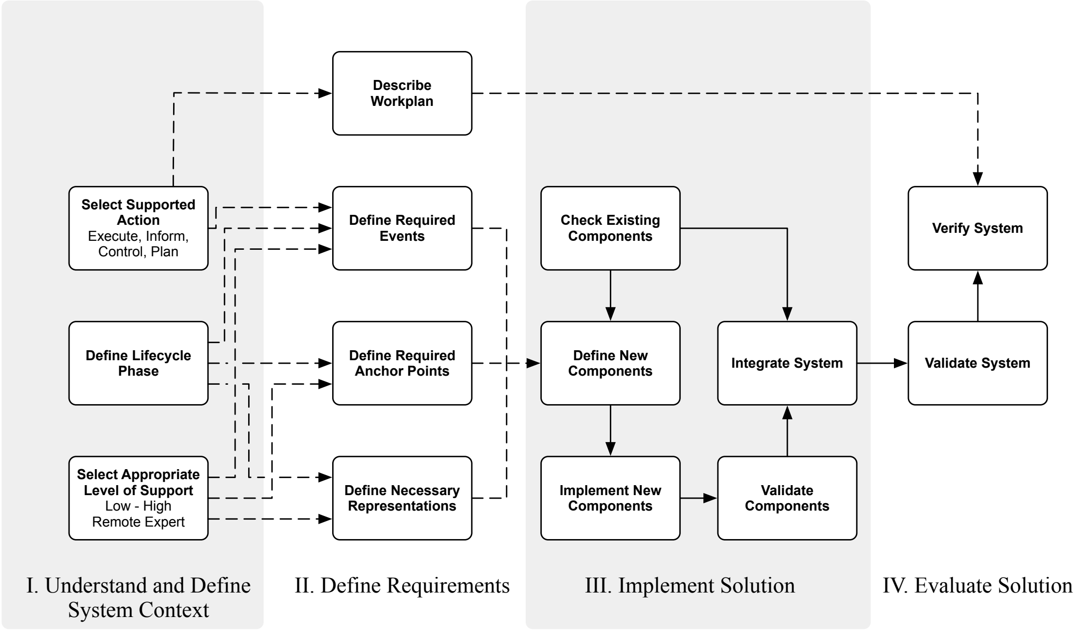


FİGURE 1. Process for Defining and Implementing a Modular IAR System

A. Understanding and Defining Context

As already stated, the usage context of an IAR application can be classified into various categories. Because the context in which the IAR system should be described without predetermining the technical implementation, the classification presented in [15] is used. Therefore, the system context is described by the supported action, the life-cycle phase, and the desired level of support.

The classification in these categories is the first indicator for the requirement definition in the second phase. The supported action can be one of four types: 1. an *execution task*, where a specific goal is to be accomplished; 2. a *control task* where the status quo is compared to a target state; 3. *planning tasks* in which the operator has to choose from an array of alternatives to establish the most effective course of action, given a set of predefined parameters; 4. and *inform tasks* where relevant data or insights are communicated to the operator, not with the immediate aim of achieving a specific result within the system, but to provide essential information that might be beneficial for future actions.

These actions effect the requirements of the system in two ways: the structure of the work plan as well as the type of events typically required to support them. First, the action has a fundamental influence on the structure of the work plan. As described, the work plan is a finite-state machine where the active state describes the presented information, while events form interaction systems are mapped to state transitions. In execution tasks, a state usually correlates with an atomic action that should be performed. When an event indicates the completion of the current task, the work plan moves to the next. Additionally, these tasks might be grouped together as to not distract experienced workers. This results in a tree-like structure of the work plan. Furthermore, the required events for such an execution task are typically limited to navigating this tree structure. Typical events are going to the next or previous step or viewing more or less details on one to navigate deeper inside the tree.

Controlling tasks are similar in that the properties to be checked and validated are often predefined in a task list. Therefore, the structure of the work plan and the required events are very similar. The main differences are that control tasks generate output. Therefore, the events that a system supporting control tasks supports must include a payload, e.g. by recording positions or measurements. Similar, planning tasks are performed to select from a choice of alternatives. The choices, for example selected products, configurations, or positions, must be recorded, as well. Therefore, the system also requires event that contain additional data. In contrast to control tasks, the structure of the work plan is not as predefined. It may contain valid configurations that must be converted into a suitable finite-state-machine before using it in a system.

Informing tasks do not generate output data directly. Rather, they are supporting the user in its decision-making. Similar to planning tasks, the required state-machine must be created beforehand out of possible states of the application, e.g. by including a state for each object an inform application should present data to.

The lifecycle phase that the system supports has the greatest impact on the specific data that can be presented, as some data is only available in later phases. While 3D models are often created early during product development, telemetry data is only available in the usage phase. It should be noted that the context might be described using multiple phases. When a product is in production in an assembly line, the product is in the phase "production" while the assembly line itself is in the "usage" phase. This makes it viable to show telemetry data, like manufacturing machine parameters or robot paths, while the actual product is assembled.

Defining the necessary level of support is crucial for defining and developing successful and helpful IAR applications. While implemented IAR prototypes are frequently presenting very detailed instructions [15], for example to support students in laboratory environments [29], research continues to show that this high level of support reduces performance of expert workers [3]. This is furthermore surprising considering preparing a complex task into a detailed and interactive AR instructions requires more work than using a limited scope for AR, like only using it to localize objects. A special case in the level of support is the remote expert. Here, the IAR application is only responsible for relaying information between the operator and the remote expert instead of giving support by itself.

B. Defining Requirements

While the previous section presents some general guidance broad requirements to describe the system context, the actual requirements depend on various factors and are mostly unique for any given system. This section presents influence factors that affect the actual requirements together with some sample questions that might be asked when detailing the requirements, including potential influences on the designed system.

Table1 1
Aspects and Guidelines that are relevant when defining requirements for IAR systems

ID	Category	Aspect	Guidelines
Q11	Task	Supported Action	Does the task require hands-free usage?
Q12		Accuracy	High Accuracy required? Consider external tracking systems
Q13		Task Complexity	How complex is the task? Which information needs to be provided to the worker while performing an action?
Q14		Spatial Information	Is orientation helpful? To which points in space should the operator be guided?
Q21	User	Experience	What are prior experiences of the user? Does he need continuous support or just a training phase? What information is relevant, what might be known to him?
Q22		Preferences	Can you offer choices, like what and how much information is needed. Consider alternatives to body-worn devices, like HMDs
Q23		Technology Level	How experienced is the user with technology, in general? Use simple and familiar interactions, where possible
Q31	Environment	Lightning	Is the environment well lid? If not, camera-based tracking systems might be unreliable
Q32		Location	Obscured or dark? Use spatial hints to guide user to the location
Q33		Spatial Stability	Is the environment changing or fixed? For fixed setups, spatial or projection-based AR might be a good alternative
Q41	Business Context	Existing Systems	Are there existing AR or worker support systems, for example pick-by-light systems that can be integrated?

Q42		Data Interfaces	With which interfaces should the system interact, for example to record data?

The first aspect that needs to be examined is the supported task. The overall supported action was already explained in the previous section. One central question when analysing the specific task is whether it requires hands-free usage. While execution tasks often greatly benefit from hands-free operation [25], it is always a trade-off with safety, ergonomic, and cost efficiency, which are also important requirements to consider [24]. The second aspect to consider is the required accuracy. Today's camera-based tracking is not capable of precise tracking. In [31], an accuracy requirement of 1 cm is mentioned as a rule-of-thumb at which precise AR hints should not be used. While [31] recommend using images or videos in such cases, external tracking systems can also be considered, as demonstrated in [13]. As AR is fundamentally a spatial form of presenting information, the most natural type of information to present is localization and orientation. AR can be efficiently used to guide workers to specific locations or draw attention to parts of a machine. Other aspects, like presenting 3D content, always comes with an added complexity during content generation and consumption. Symbols and texts are the most common representation of data in IAR applications, while 3D models and animations are only relevant to specific applications [16], like marketing or product design.

The second aspect to consider is the users and workers that will use the assistant system. Central questions are their general experience with the specific process, they individual preferences, but also the technology level they are familiar with. First, experienced workers need different support and information than novice workers or employees with cognitive impairments. Showing to many or to detailed instructions might increase the cognitive load for experienced workers, while novice workers want reduced information as they get more proficient with the task at hand [3]. For novice workers, the usage of AR has shown to reduce the training time and number of training cycles when performing complex tasks at the first time [32]. Therefore, one should consider using AR for training of workers not experienced with the specific tasks, unless variations or other requirements require continuous access to information for experts as well. On the other hand, workers with cognitive impairments might benefit from detailed instructions that are visible at all times [33]. Overall, one should always question whether data representation is actually necessary in this form for the worker at a specific task. Instead of using sophisticated hardware to display 3D animations, the necessary spatial information might be provided using a traditional pick-by-light setup, instead.

As different information can be presented in different ways using AR [34], one can also consider the preference of the users when designing an IAR system. This might include placement of information, filtering and deciding on the level of support wanted, to basic accessibility settings like text sizes. Similarly, the interactions with such a system should be designed multimodal [6], for example by adding an alternative input method to voice and dictation. Similarly, the technology level the users are familiar with need to be considered. Studies show that HMDs in particular tend to have higher mental load than other AR and non-AR devices, like projection-based or mobile AR or paper manuals [35]. Mobile AR, for example, uses smartphones and tablets with interactions the users might be more familiar than novel HMDs. This also underlines that the most sophisticated IAR application might be unsuitable for a given task if it does not take into considerations the needs and requirements of the user.

The physical environment in which the system should be used limits the technologies that can be implemented successfully. In dark environments, camera-based AR tracking is often limited. At the same time, users benefit from spatial information and directions in these circumstances [31]. When the environment and setup is not changing, fixed AR setups offer better accuracy and reliability, for example by using external tracking systems. Data can also be made available through projection-based AR, which reduces mental load and improves economy, which are important aspects to consider when defining the requirements.

Finally, the business context needs to be taken into account. Important requirements are the existing business processes in order to define the data that the system needs to collect and the data format used. This could, for example, be used to create a maintenance report after an execution task or saving results of quality control. Then existing interfaces to other enterprise tools, like product lifecycle management systems for product data and models, enterprise resource planning for tasks and orders, as well as internet of things platforms for live telemetry. For these interfaces, data availability and formats, potential conversions as well as access rights and caching potentials.

Overall, the given aspects and considerations can be used as a framework when defining requirements for IAR systems. They are developed based on experience and best practices from the literature. These general considerations can offer a good baseline from which to specify more detailed requirements that fit the use case to implement.

C. Implement Solution

Based on the requirements that have been defined in the previous step with help from the provided guidelines, the next step is to implement the IAR system. Then, suitable components have to be defined or selected by breaking down and decoupling the functionality required by the overall system. These include the required interactions and events, the data that should be presented, and the spatial information that is necessary for the user.

As reusability is one of the main advantages of the proposed architecture, checking and evaluating already existing components is a first step. As each component is defined by supported events, data representations, and anchor points, these can be evaluated against the requirements defined before. When deciding on reusing components, one can consider transforming data into a simpler representation that is supported by the existing one, for example by rendering a 3D model into a 2D rendering or converting it into supported formats. In general, the overall requirements can be fulfilled by combining components. Providing both assembly instructions and supporting the localization of parts to be used during an assembly step might be implemented by a pick-to-light component for the parts and a monitor component that displays the instructions. Alternatively, both features can be integrated into a single component, for example in an HMD or using projection-based AR.

When the requirements cannot be satisfied using existing components, new components need to be implemented. First, the required events, anchors, and representations need to be split into one or more components to implement. In general, one should separate interaction components and their events from presentation components that support anchors and representations. Interaction components often support multiple events. A voice recognition system might enable event to navigate through a work plan by proceeding to the next or previous event or asking for help. Exceptions are events that hold additional data. A component where the operator enters information, for example the results of a conducted measurement, typically only support a single event type that holds the entered information. For presentation components, the separation should be done primarily by their supported anchors. These describe which objects or other features the component can recognize and use to display information near them. Typically, a presentation component only supports a limited number of different objects. They are often only available after the component is calibrated. This could include providing reference images or 3D models that are used in the tracking algorithms.

In the next phase, the components are implemented individually based on their defined interfaces. This decomposition enables the work to be split across teams or companies. For example, a team could implement a model tracking algorithm and provide it as a component without being concerned with how this functionality can be used in an IAR system. This separation of concern greatly increases the flexibility and efficiency in the implementation phase The component is then described by its capabilities. For an interaction component, these are the events that the component emits, while presentation components are described by supported anchor points and data representations. Additionally, each component is assigned a unique identifier. In this phase, the implemented component is verified against their requirements to ensure that it is able to fulfil the advertised capabilities.

Both existing and newly implemented components are then integrated into the overall system. The application planning module collects the individual capabilities and compares them to the required capabilities from the work plan that should be executed. From there, it can be determined if the overall system can support all features required. As supported anchors might require additional calibrated, for example by providing tracking information for additional objects, at this stage, the planning module might prompt the operator to configure missing capabilities. Only after all capabilities are fulfilled can the system be used with the provided workplan.

D. Evaluate Solution

Finally, the overall system needs to be evaluated. Common approaches are validation first in a laboratory environment and then in the final environment by end users [28]. The NASA TLX index is often used to evaluate the performance of IAR systems, sometimes in combination with performance indicators, like learning time, task execution time, and error rate [29]. Overall, it is important to evaluate a solution in the specific workplace where it will later be used to identify problems with the environment, such as camera-based tracking issues or speech recognition. Additionally, workers who will later use the system, need to test the implementation in this phase, as well. During this phase, issued surfaced, and general feedback can be used to further improve individual components and the overall system. Finally, it is important to check the reusability of the newly introduced components to ensure that they can also be used in future systems.

IV. CASE STUDY

To demonstrate the proposed methodology, a case study with an assembly support system is presented in this section. After describing the context of use, the requirements are defined on the basis of the key questions outlined above. Finally, it is described how the system is decomposed and implemented from individual components.

A. Understanding and Defining Context

The system to be implemented is an assembly support system for the assembly of toy robots. The robots are about 20 cm long and are assembled using plates and screws. The required parts are in storage containers at the assembly station, while a simple fixture is used during the process. The assembly stand is used in a laboratory environment mostly by students and guests without deeper experiences with the assembly. The desired degree of support can be considered medium to high, as detailed and ongoing descriptions of tasks to be performed should be shown to the user. However, different users will have different skills and experiences. The required level of support implies a structured work plan. This can be generated based on the assembly instruction, for example by following the process outlined in [23]. The resulting tree-like data structure describes an assembly step in each state, while similar steps are grouped together to account for the expected varying skill levels of the users.

In the lifecycle phase production, full 3D models of the overall assembly as well of the assembly steps can be provided. Textual descriptions for the steps can be provided. The necessary parts required in each assembly step can be extracted from the assembly procedure, as well. Due to the unknown assembly state, the support system should provide guidance so that the operators can quickly pick the correct part for each assembly step.

B. Defining Requirements

Following the outlined guidelines in the previous chapter, the requirements for the overall system can be defined. The supported action (Q11) is an execution task, making hands-free usage an important consideration. In fact, this use-cases benefits a lot from it, because the operator will regularly use both hands, e.g. when screwing the plates together. Hands free usage is therefore desired. However, it is not complex enough to warrant 3D animations (Q13).

As only instructions and 3D models are used, there is no need for especially high tracking accuracy (Q12). Spatial information can be used to guide the user to the correct storage container and highlight the points on the assembly, e.g. for showing screws, as operators are not familiar with the assembly stand (Q14)

As the support system will be used mainly as a demonstrator there will be no dedicated training phase. Instead, users will require ongoing support, but this will vary depending on their previous experience (Q21). As this is purely a demonstrator, the system does not need to adapt to preferences of the user (Q22). The technology level the users are used to is expected to be high (Q23), so a full-fledged IAR application including advanced features, like voice or gesture recognition, can be used. The environment is a laboratory setting, so lightning (Q31) and obstructions (Q32) will not be a concern. Because of the dedicated assembly stand, including fixtures, there is a high spatial stability (Q33). The setup and surroundings will not change often. This makes a fixed setup, e.g. by using a projector, a viable option. Additionally, the setup can be calibrated manually, for example by manually configuring spatial hints. This allows the usage of object anchors without dedicated object tracking capabilities. The locations of the storage containers or screws, for example, can be calibrated and saved once or after changes in the setup. While data and information is available in enterprise systems, like 3D models in a PLM environment (Q41), manually preparing and converting data for the demonstrator is a viable alternative to implementing data interfaces (Q42). Based on these requirements and concerns, the system can be decomposed into individual components.

C. Implement Solution

The main requirements for the system are presentation of textual instructions, a 3D model of the current assembly step, symbols or indicators at objects or parts of objects, interactions for navigating through the state machine, and recognition of the parts the user has taken out of storage containers. Therefore, the system requires text, symbol, and 3D model data representations, as well as interactions for navigation (next, previous, and help), and the selection of specific parts. This can be broken into individual components as shown in figure 3. The components are implemented in two dedicated modules. The first is responsible for rendering the instructions and the 3D model, together with the navigation interactions. The second is capable of indicating the position of storage container and can register when the user grabs a part out of one of them.

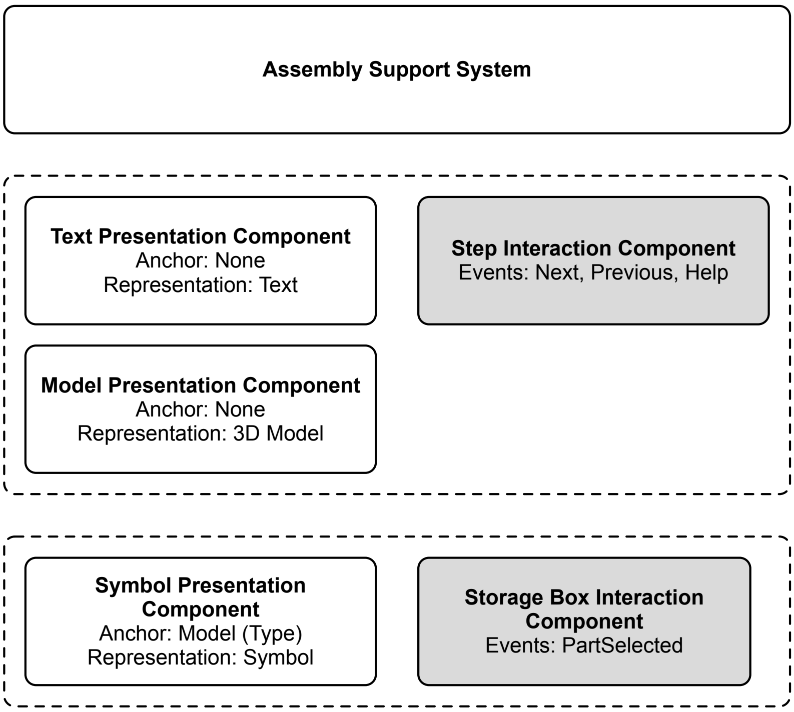


FİGURE 2. Implemented Components for the Assembly Support System

Overall, the assembly station is equipped with a projector and a depth sensor. Instructions and images are presented directly on the work surface. The position of the storage container is set manually. The depth sensor recognizes gestures as well as the grabbing of objects out of those. The implemented system can be seen in figure 4.

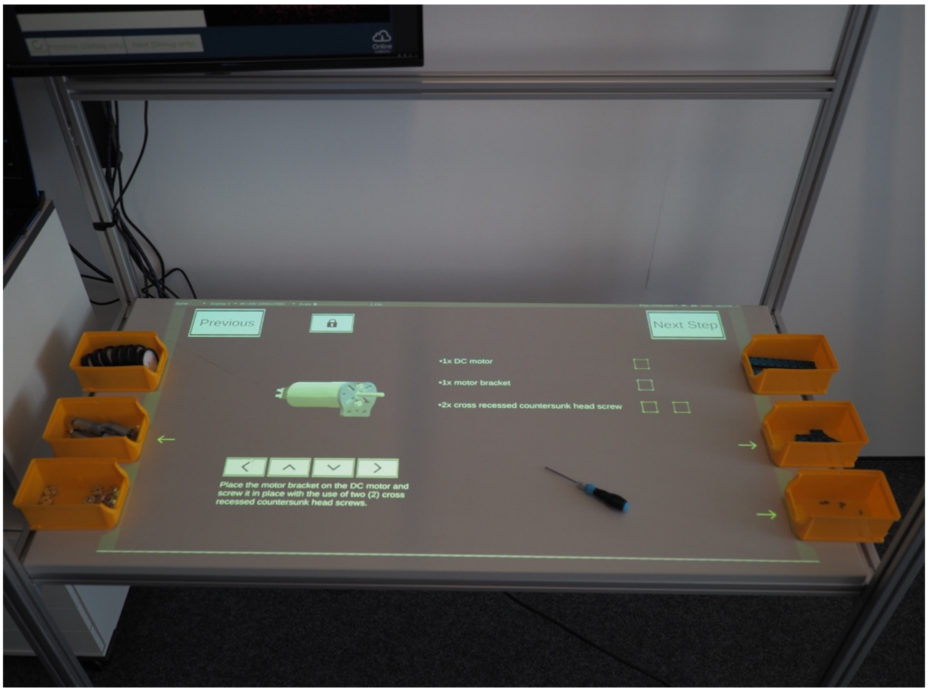


FİGURE 1. The implemented projection stand with highlighted storage containers, instructions, and the 3D model.

D. Evaluate Solution

The individual components have been verified to ensure they fulfil the given requirements. Because the system will be used exclusively for demonstrating purposes, an evaluation and formal validation has not been performed. However, before implementing such a system in a production environment, evaluation should be performed with the target group to ensure the system is supporting them adequately.

V. CONCLUSION

To reduce effort required when introducing and implementing IAR applications in the industry a modular approach offers various advantages. Individual components can be implemented by specialists and reused in other systems. To better fulfil special requirements of individual users, by integrating new components into an existing system. However, such a modular approach increases the complexity when designing a system. In addition to requirements for the overall system, it needs to be decomposed into individual modules and components. A system designer needs to select the functionality for each component so that the overall system can fulfil the requirements, while the individual components are general enough to make them reusable.

This contribution presents a methodology for defining requirements for such a component-based architecture. Based on established standards and best-practices from the literature, a four-step process is described. Guideline questions are provided that support in the requirement definition phase of IAR systems. While the provided case-study focuses on a laboratory setting, the generated insights can be transferred to other use-cases, as well.

ACKNOWLEDGMENT

The headings of the Acknowledgment and the References sections must **not** be numbered. Editorial Board wishes to acknowledge all authors for their kind attentions in complying with the formatting instructions in this template.

REFERENCES

[1] P. Fite-Georgel, "Is there a reality in industrial augmented reality?" in *International symposium on mixed and augmented reality*, in Science and technology proceedings. IEEE, Oct. 2011.

[2] S. Büttner, M. Prilla, and C. Röcker, "Augmented reality training for industrial assembly work - are projection-based AR assistive systems an appropriate tool for assembly training?" in *2020 CHI conference on human factors in computing systems*, New York, NY, USA: Association for Computing Machinery, 2020, pp. 1–12. doi: 10.1145/3313831.3376720.

[3] M. Funk, A. Bächler, L. Bächler, T. Kosch, T. Heidenreich, and A. Schmidt, "Working with augmented reality?" in *10th international conference on PErvasive technologies related to assistive environments*, New York, NY, USA: Association for Computing Machinery, Jun. 2017, pp. 222–229.

[4] E. Bottani and G. Vignali, "Augmented reality technology in the manufacturing industry: A review of the last decade," *IISE Transactions*, vol. 51, no. 3, pp. 284–310, Feb. 2019.

[5] V. Bilous, R. Porsch, and K. Spanoudakis, "Augmented reality: Increasing availability and its implication for SMEs," *The Future of Smart Production for SMEs*, pp. 349–356, Oct. 2022, doi: 10.1007/978-3-031-15428-7_31.

[6] W. Fang, L. Chen, T. Zhang, C. Chen, Z. Teng, and L. Wang, "Head-mounted display augmented reality in manufacturing: A systematic review," *Robotics and Computer-Integrated Manufacturing*, vol. 83, p. 102567, Oct. 2023, doi: 10.1016/j.rcim.2023.102567.

[7] T. Masood and J. Egger, "Adopting augmented reality in the age of industrial digitalisation," *Computers in Industry*, vol. 115, p. 103112, 2020.

[8] "Ergonomie der mensch-system-interaktion – teil 210: Menschzentrierte gestaltung interaktiver systeme," Deutsches Instut für Normung e.V., DIN EN ISO 9241-210, 2020.

[9] M. Runji, Y.-J. Lee, and C.-H. Chu, "Systematic literature review on augmented reality-based maintenance applications in manufacturing centered on operator needs," *International Journal of Precision Engineering and Manufacturing-Green Technology*, vol. 10, no. 2, pp. 567–585, Mar. 2023, doi: 10.1007/s40684-022-00444-w.

[10] T. Kosch, Y. Abdelrahman, M. Funk, and A. Schmidt, "One size does not fit all," *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, Sep. 2017, doi: 10.1145/3123024.3124395.

[11] A. Syberfeldt, M. Holm, O. Danielsson, L. Wang, and R. L. Brewster, "Support systems on the industrial shop-floors of the future – operators' perspective on augmented reality," *Procedia CIRP*, vol. 44, pp. 108–113, 2016.

[12] M. Eswaran, A. K. Gulivindala, A. K. Inkulu, and M. V. A. Raju Bahubalendruni, "Augmented reality-based guidance in product assembly and maintenance/repair perspective: A state of the art review on challenges and opportunities," *Expert Systems with Applications*, vol. 213, p. 118983, Mar. 2023, doi: 10.1016/j.eswa.2022.118983.

[13] G. Stübl, G. Ebenhofer, H. Bauer, and A. Pichler, "Lessons learned from industrial augmented reality applications," *Procedia Computer Science*, vol. 200, pp. 1218–1226, 2022, doi: 10.1016/j.procs.2022.01.322.

[14] D. Röltgen and R. Dumitrescu, "Classification of industrial augmented reality use cases," *Procedia CIRP*, vol. 91, pp. 93–100, 2020, doi: 10.1016/j.procir.2020.01.137.

[15] J. L. Siewert, M. Neges, and D. Gerhard, "Ein klassifizierungssystem für industrielle augmented reality anwendungen," pp. 401–416.

[16] M. Bauer *et al.*, "Design of a component-based augmented reality framework," in *Proceedings IEEE and ACM international symposium on augmented reality*, in ISAR-01. IEEE Comput. Soc, 2001. doi: 10.1109/isar.2001.970514.

[17] M. Schumann, P. Klimant, C. Kollatsch, and V. Wittstock, "Modulares framework zur entwicklung von augmented-reality-anwendungen," Feb. 2013.

[18] P. Fraga-Lamas, T. M. Fernandez-Carames, O. Blanco-Novoa, and M. A. Vilar-Montesinos, "A review on industrial augmented reality systems for the industry 4.0 shipyard," *IEEE Access*, vol. 6, pp. 13358–13375, 2018.

[19] T. Kuster *et al.*, "A distributed architecture for modular and dynamic augmented reality processes," in *2019 IEEE 17th international conference on industrial informatics (INDIN)*, IEEE, Jul. 2019. doi: 10.1109/indin41052.2019.8972101.

[20] E. Laviola, M. Gattullo, A. Evangelista, M. Fiorentino, and A. E. Uva, "In-situ or side-by-side? A user study on augmented reality maintenance instructions in blind areas," *Computers in Industry*, vol. 144, p. 103795, Jan. 2023, doi: 10.1016/j.compind.2022.103795.

[21] M. Gattullo, A. Evangelista, A. E. Uva, M. Fiorentino, and J. L. Gabbard, "What, how, and why are visual assets used in industrial augmented reality? A systematic review and classification in maintenance, assembly, and training (from 1997 to 2019)," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 2, pp. 1443–1456, Feb. 2022, doi: 10.1109/tvcg.2020.3014614.

[22] D. Gerhard, M. Neges, J. L. Siewert, and M. Wolf, "Towards universal industrial augmented reality: Implementing a modular IAR system to support assembly processes," *Multimodal Technologies and Interaction*, vol. 7, no. 7, 2023, doi: 10.3390/mti7070065.

[23] M. Quandt, A. A. Alla, L. Meyer, and M. Freitag, “Success factors for the development of augmented reality-based assistance systems for maintenance services,” in *7. WGP-jahreskongress*, R. H. Schmitt and G. Schuh, Eds., Apprimus Verlag, 2017.

[24] M. Quandt, B. Knoke, C. Gorldt, M. Freitag, and K.-D. Thoben, “General requirements for industrial augmented reality applications,” *Procedia CIRP*, vol. 72, pp. 1130–1135, 2018, doi: 10.1016/j.procir.2018.03.061.

[25] M. Lorenz, S. Knopp, and P. Klimant, “Industrial augmented reality: Requirements for an augmented reality maintenance worker support system,” in *Adjunct proceedings of the 2018 ieee international symposium on mixed and augmented reality (ismar)*, 2018, pp. 151–153. doi: 10.1109/ISMAR-Adjunct.2018.00055.

[26] S. Aquino, M. Rapaccini, F. Adrodegari, and G. Pezzotta, “Augmented reality for industrial services provision: The factors influencing a successful adoption in manufacturing companies,” *Journal of Manufacturing Technology Management*, vol. 34, no. 4, pp. 601–620, Feb. 2023, doi: 10.1108/jmtm-02-2022-0077.

[27] M. Gattullo, G. W. Scurati, M. Fiorentino, A. E. Uva, F. Ferrise, and M. Bordegoni, “Towards augmented reality manuals for industry 4.0: A methodology,” *Robotics and Computer-Integrated Manufacturing*, vol. 56, pp. 276–286, Apr. 2019, doi: 10.1016/j.rcim.2018.10.001.

[28] M. Quandt, T. Beinke, and M. Freitag, “User-centered evaluation of an augmented reality-based assistance system for maintenance,” *Procedia CIRP*, vol. 93, pp. 921–926, 2020, doi: 10.1016/j.procir.2020.03.053.

[29] A. Dey, M. Billinghurst, R. W. Lindeman, and J. E. Swan, “A systematic review of 10 years of augmented reality usability studies: 2005 to 2014,” *Frontiers in Robotics and AI*, vol. 5, Apr. 2018, doi: 10.3389/frobt.2018.00037.

[30] A. Syberfeldt, O. Danielsson, and P. Gustavsson, “Augmented reality smart glasses in the smart factory: Product evaluation guidelines and review of available products,” *IEEE Access*, vol. 5, pp. 9118–9130, 2017, doi: 10.1109/access.2017.2703952.

[31] K. Tainaka, Y. Fujimoto, T. Sawabe, M. Kanbara, and H. Kato, “Selection framework of visualization methods in designing AR industrial task-support systems,” *Computers in Industry*, vol. 145, p. 103828, Feb. 2023, doi: 10.1016/j.compind.2022.103828.

[32] T. Bosch, G. van Rhijn, F. Krause, R. Könemann, E. S. Wilschut, and M. de Looze, “Spatial augmented reality,” in *Proceedings of the 13th ACM international conference on PErvasive technologies related to assistive environments*, New York, NY, USA: Association for Computing Machinery, Jun. 2020. doi: 10.1145/3389189.3397975.

[33] M. Jost, A. Luxenburger, S. Knoch, and J. Alexandersson, “PARTAS: A personalizable augmented reality based task adaption system for workers with cognitive disabilities,” *Proceedings of the 15th International Conference on PErvasive Technologies Related to Assistive Environments*, Jun. 2022, doi: 10.1145/3529190.3529208.

[34] M. Gattullo, G. W. Scurati, A. Evangelista, F. Ferrise, M. Fiorentino, and A. E. Uva, “Informing the use of visual assets in industrial augmented reality,” 2019, pp. 106–117. doi: 10.1007/978-3-030-31154-4 10.

[35] M. Funk, T. Kosch, and A. Schmidt, “Interactive worker assistance,” *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, Sep. 2016.

International Journal of Advances in Production Research

Insights from Dynamic Pricing Scenarios for Multiple-generation Product Lines with an Agent-based Model using Text Mining and Sentiment Analysis

Atefeh Anisi[a,1], Gül E. Okudan Kremer[b,2], Sigurdur Olafsson[a,3*]

[a] *Department of Industrial and Manufacturing Systems Engineering, 3004 Black Engineering, Iowa State University, USA*

[b] *College of Engineering, University of Dayton, USA*

[1]anisi@iastate.edu. [2]gkremer2@udayton.edu. [3]olafsson@iastate.edu.

Abstract— **Corporations must constantly upgrade and improve their offerings due to changes in customer preferences. It is a common strategy for firms in technology-intensive markets to use online reviews as a source of product information to inform such changes. This user-generated information is valuable since it provides companies with valuable and low-cost input. In this paper, we propose an agent-based model for simulating potential cannibalization situations with respect to customer satisfaction throughout consecutive generations of a product line. The level of customer satisfaction is regarded as a parameter in the model, which is conceptualized to affect the product price. The proposed model provides insights into different pricing strategies regarding customer satisfaction levels affect the total lifecycle profitability of multiple-generation product lines, and how they can be used to assist organizations in developing appropriate dynamic pricing strategies.**

Keywords— **text mining, sentiment analysis, multiple-generation product lines, cannibalization, agent-based modeling, dynamic pricing scenarios.**

I. INTRODUCTION

We are surrounded by multiple-generation products (MGP), from simple ones like NIVEA skin care products to the most complicated ones like Apple Mac Books. In MGP lines, a company will introduce the first version of a product to the market and periodically introduce the following versions. Different versions of an MGP have the same core functionality, each with higher technologies or more features than the previous versions [1]. Choosing multiple-generation product strategies has advantages for businesses but may also create cannibalization. Market cannibalization is a loss in sales, revenue, or market share of a company caused by introducing a new product similar to another product in the market. For MGPs, cannibalization refers to the competition that may happen between different generations of a product. When the company releases its new generation, the price of previous generations may be discounted, and therefore, they may attract customers due to their lower price. This, in turn, leads to less profit for the company since the latest generation typically has the highest profit margin.

Despite the risk of cannibalization, research shows that having a multiple-generation product line can be 26% to 40% more cost-effective than introducing a single generation of a product or sequentially introducing a multiple-generation product [2]. Choosing multiple-generation product strategies can benefit companies in different aspects. For example, with multiple-generation product line thinking, companies should have a longer product lifespan to allow more time to develop new products, and therefore, they can use resources better, keep their market share, and finally earn an optimal profit level in the long run [3].

Multiple-generation products differ from product families where the company produces different products for the customers with different priorities to cover all market segments. In contrast, multiple-generation products are introduced to the market over time, and all versions aim to cover the same market segment [4,5]. For example, Apple iPhone 12 series are different generations in a product line; although different, they are all designed for the same market niche and target the same customer group. Their core functionality is that they are all high-performance cell phones. In product families, the same product will be offered in different sizes, flavors, colors, and textures, but always the same product. The goal is to provide various products to the market for different market niches.

---

[*] Corresponding Author

In prior literature, several approaches were used to model the long-term behavior of multiple-generation products, including the Dynamic State Variable Model (DSVM). DSVM can be successfully used for MGPs and effectively forecast the sales performance and the introduction timing of each generation regarding the cannibalization. In their related work, Lin and Okudan applied DSVM using the Apple Inc. iPhone product line with its sales data [3,6]. Lin and Okudan applied DSVM to generate full performance predictions for a new multiple-generation product line, namely Apple iPads, using the historical data from a similar product line if Apple iPhones [6].

To cope with cannibalization, companies need to choose the best pricing strategy to balance the sales from different generations present in the market at each period. Accordingly, the company needs a model that can simulate the cannibalization scenarios within a product line with multiple generations and generate pricing strategies. Lin, Kilicay-Ergin, and Okudan proposed an agent-based model (ABM) that can help companies choose the best pricing strategies to maximize their profit in the long run [7]. They used this model along with a DSVM based on a two-phase methodology. In the first phase, the sales forecast and introduction timings are acquired through the DSVM model. In the second phase, the DSVM outputs are used as the input data for the agent-based model to generate optimal pricing strategies during the product life cycle. Kilicay-Ergin, Lin, and Okudan used their suggested two-phase methodology for Apple Inc.'s iPhone line [8].

With the technology improvement, online reviews are one of the best sources of information on products. This user-generated information is valuable for manufacturers because they can give companies vital feedback for free while achieving this information through surveys or focus groups can cost companies. Companies need to continuously update their decisions and strategies because of the continuous changes in customer preferences to make items attractive or remove product deficiencies.

Parallel simulation models, which can evaluate all the distributed pieces and their complicated interactions concurrently, have numerous advantages over traditional modeling tools for large adaptive systems with multiple distributed interacting parts. Particularly, the agent-based model is a framework capable of simulating these parts (agents), their decisions, their interactions, and the system's overall behavior. The ABM proposed herein considers MGPL as a complex adaptive system and analyzes the whole system's behavior, enabling analysis of pricing strategies [8]. MGPL is considered a complex adaptive system since each agent (generation) in this system is authorized to change its price and adapt to the market considering the current market situation. Each product generation's strategic decisions influence the overall system behavior in this system.

Natural language processing (NLP) enables computers to automatically comprehend text and words as humans do, and sentiment analysis is the use of NLP to identify, extract and study subjective information. Sentiment analysis can be widely applied to analyze reviews or surveys. There is prior research that uses text mining techniques to evaluate customer satisfaction towards products from relevant reviews. There are also papers that use ABM to model MGP pricing scenarios. The main contribution of the present work is to use text mining techniques and pricing scenarios simultaneously to generate appropriate pricing strategies that inherently take customers' opinions into account. We extend the work in [8] by considering the consumer's opinion towards different product generations as a factor using text mining techniques on the same case study (Apple iPhones). Although there are numerous related works, to our knowledge, text mining techniques have never been used in combination with ABMs. In this work, we propose a two-phase framework based on the ABM model to analyze pricing scenarios for companies. In the first phase, we apply text mining and sentiment analysis techniques to extract user reviews from the Apple community forum and evaluate the consumer's satisfaction level for each product generation in each time-period. In the second phase, we use the outputs from the previous step in the agent-based model when generating the best pricing strategies during the product lifecycle. Figure 1 shows these steps respectively.

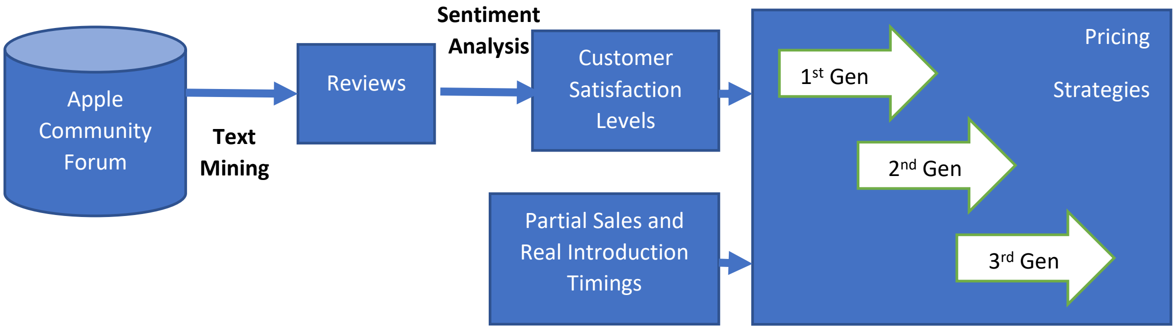


FIGURE 1. OVERVIEW OF AGENT-BASED DYNAMIC PRICING MODEL FOR MULTI-GENERATION PRODUCT LINES (MGPL) WITH THREE GENERATION

II. LITERATURE REVIEW

A. Multiple Generation Product Lines and Quantitative Models to Analyze Their Behavior

Due to today's technology-intensive and rapidly changing market environments, applying an MGP strategy becomes a favorable tool for companies. For example, Apple Inc. became one of the most profitable companies in the world with its three well-known MGP lines: The Apple iPods, iPhones, and iPads. Ofek and Sarvary studied the competition between market leaders and followers, created a multi-period Markov game model, and evaluated the effects of innovative advantage, reputation advantage, and advertising on R&D for leaders through their model [9]. Moreover, they investigated appropriate strategies for followers. They discovered that the market leader concentrates on strong demand for its existing product by investing more in advertising, whereas the follower focuses more on R&D. Their analysis gives useful insights for developing a marketing strategy in fast-paced, high-tech business environments. Microsoft (MS) focused its R&D on successive generations of products and applied forward-looking MGP strategies on all its product lines. For instance, when MS launched its Windows 7, its developers had the idea of Windows 8 in their minds. MS looks back on its strategies periodically to review and adjust them if needed [10]. The quantitative models used for multiple-generation products can be categorized into behavioral, dynamic competition, and pricing models. Behavioral models try to interpret the sales behavior of the product during its lifecycle, while dynamic competition models consider the market as a competitive environment and try to generate competitive scenarios for companies. Pricing models tend to create pricing policies and determine optimal pricing strategies for companies. Norton and Bass utilized the bass diffusion model to analyze the MGP sales behavior with respect to the substitution effect among consecutive generations to predict the changes that may happen to future demands of MGPs [11]. In the substitution effect, a portion of the demands from the current product generation is replaced with successive ones. Mahajan and Muller expanded this work by proposing a new demand model that considers the adoption and substitution effects, not only between two consecutive generations but also across different generations, named the "leapfrog" effect [12]. Their model is capable of generating optimal timing strategies regarding cannibalization. Morgan et al. evaluated the quality versus time-to-market trade-offs for MGPs [2]. They proposed an innovative model regarding multiple-generation product lines with different factors, including additional costs, the focal firm's quality, competitive quality, and market share with an active competitor to maximize the total profit.

Krankel et al. used a dynamic programming technique to model MGPs [1]. They proposed a multi-stage decision model to investigate successive introduction timing strategies and forecast future demands while the technology level is additive, and the new generation of the product completely replaces the previous one. Bardhan and Chanda proposed a new behavioral model incorporating the Bass diffusion model and extended [12] by dividing the cumulative adaptors into two categories: First-time purchasers and repeat purchasers and modeling them, respectively [13]. Huang and Tzeng suggested a new two-stage fuzzy piecewise regression for predicting product lifetime and annual MGP shipments [14]. First, the lifetime of each product will be predicted based on the historical data through the proposed regression model, and then the annual shipments of each product will be determined.

Dobson and Kalish developed a tool to help managers in the process of product line design and pricing by providing three types of information: (1) what kind of product customers want, (2) the cost to produce each type of product, and (3) information about their current and future products [15]. They determined the introduction timings and pricing strategies using a heuristic algorithm to maximize the profit. Arslan et al. proposed solutions for firms to manage introducing and pricing their product generations in the competitive market [16]. They analyzed both monopolistic and duopolistic (competitive) environments. They provided introduction timings and optimal pricing policies for two successive product generations under complete replacement and when there are coexisting generations. They proposed a competition model between two firms under complete replacement.

Some pricing models for MGPs are based on the multinomial logit (MNL) models. MNL models assume that the probability of a customer's purchase depends on the customer's utility function for each product [17]. These models perform admirably when the products are distinct from one another. Nonetheless, a number of studies demonstrate the disadvantages of the MNL models when there is a link between product alternatives [17,18]. Kim et al. created a purchase timing and generation choice model for initial and recurring purchases of multiple-generation products [19]. Their model has a logit formulation and models the likelihood of purchasing a generation given prior purchases. Their approach uses individual purchase histories to explain repeating purchase behavior. Schön considered an MNL model with price discrimination among customers [20]. The authors used probabilistic customer choice models to make pricing decisions for MGPs.

Pricing is an effective tool for companies to prevent or at least mitigate problems during an inter-generational product transition since uncertainty in the introduction of a new product may lead to mismatches between demand and supply. The authors considered a transition in which a new generation product replaces an old one, assuming the new one has better features and performance than the old one. They formulated a dynamic pricing problem and derived the optimal pricing policies for both the old and new product generations. In addition to product replacement, they considered several dynamics, such as substitution, external competition, scarcity, and inventory, as well as how these factors affect pricing policies. They also determined the optimal initial inventory for each product [17]. Chen and Chang suggested a dynamic programming model to manage new and remanufactured products by studying their pricing behavior over their life cycle length [21]. The authors considered the new product and the remanufactured product as subsequent generations. The primary purpose of their formulation is to examine the pricing behavior of the product under different parameter settings such as manufacturing and remanufacturing costs, market growth rate, return rate, and substitutability. Fruchter et al. employed genetic algorithms, a mathematical heuristic mimicking the process of biological evolution, to the problem of optimal product line design to generate pricing decisions [22]. Special operators were used to help genetic algorithms mitigate cannibalization. The authors considered the manufacturer's profit as the criteria for fitness-evaluating chromosomes.

Agent-based models (ABM) consist of entities (agents) and a framework to simulate agent decisions and interactions. In a multiple-generation product line, each generation acts as an independent agent and will adjust its sales price according to market demand. These agents' decisions and interactions generate the general behavior of the system. The advantage of using an Agent-based model compared to MNL is that ABM lets decision-makers analyze cannibalization scenarios over the MGP life cycle, while it is challenging to study cannibalization scenarios using MNL since there is a correlation between different generations [8]. Also, the ABM lets the decision-makers understand the implications of varying pricing decisions on the firm's overall profit. We use ABM as the core of our two-phase framework to evaluate different pricing scenarios for Apple iPhones with respect to the importance of consumers' feedback. We need sentiment analysis techniques to gather, integrate, and interpret users' reviews; then, we will use the information extracted from them in our framework.

B. Text Mining and Sentiment Analysis

In the last few decades, a huge amount of information has been generated in text format. Text mining refers to the process of extracting knowledge and information from unstructured text [23]. Zhan et al. analyzed the consumers' reviews using text mining to extract the customers' concerns and summarize topics based on their rankings using an automatic text summarization approach. [24]. The authors compared their method with other approaches, such as opinion mining. Thorleuchter et al. used text mining and text classification (tokenization, term filtering method, Euclidean distance measure, etc.) techniques and a novel heuristic measure for idea mining [25]. The process of extracting new and useful ideas from unstructured text data is known as idea mining. Their evaluated approach is implemented as a web-based application titled "Technological Idea Miner".

Sentiment Analysis (SA) is a method to identify and categorize opinions expressed in a text to determine the sentiment behind it. Using fundamental sentiment analysis, a program can determine whether a text's sentiment is positive, negative, or neutral regarding a special product or service [26]. Specifically, it is a process of analyzing people's opinions and emotions in some special piece of text. There are two primary applications of sentiment analysis [27]. First, sentiment analysis has been applied to documents to differentiate between positive and negative reviews [28-31]. Second, it has been applied at the sentiment level to accomplish some tasks such as multi-perspective question answering and summarization, opinion-oriented information extraction, and customer review mining [27, 32-35].

Sentiment analysis starts with sentiment expressions existing in the specified object and then recognizes positive from negative words and phrases [36]. Lexicons can be categorized into three types: Positive polarity (e.g., excellent, great, perfect), negative polarity (e.g., bad, terrible, awful), and contextual polarity (i.e., words with different meanings in different contexts). Turney used a simple method to classify reviews into two types: recommended and not recommended, according to their average semantic orientation value of phrases containing adjectives or adverbs [31]. Pang and Li used a machine learning technique to apply text categorization methods to the subjective part of any document based on the minimum number of cuts [29]. Beineke et al. considered the traditional sentiment classification method but as a Naïve-Bayes model [28].

Some other works focused on sentiment analysis at the sentence or phrase level. Wilson et al. proposed a new approach to classify the expressions into neutral or polar and disambiguate the polar expressions at the phrase level [27]. Recently, Täckström and McDonald developed two semi-supervised latent variable models for sentiment analysis at the sentence level [37]. Mostafa used text mining techniques on 3516 tweets during a specific time period to evaluate the consumers' sentiment towards brands such as Nokia, T-Mobile, IBM, KLM, and DHL [38]. The author showed that there is positive sentiment toward some famous brands. Kontopoulos et al. suggested an ontology-based method for sentiment analysis of Twitter posts [39]. In their approach, each distinct notion in a post will receive a sentiment grade instead of assigning a total sentiment score to the post. This can finally lead to a more detailed analysis of post opinions regarding a specific topic. There are some works about the details of the sentiment analysis approach. Deng et al. introduced a strategy that gives terms specific weights to improve sentiment analysis performance [40]. They proposed a supervised term weighting scheme based on two main factors: The importance of a term in a document (ITD) and the importance of a term for expressing sentiment (ITS). The authors introduced seven statistical functions that learn the sentiment importance of a term through its statistical distribution in positive and negative documents.

Although the studies mentioned above focused on sentiment analysis at document and sentence levels, they are only about techniques, not their usage in combination with different frameworks. These studies cannot derive useful information about the usage of Text mining for customer satisfaction evaluation. Kang and Park proposed a sentiment-analysis-based framework for measuring customer satisfaction using the VIKOR approach [42]. They applied their framework to customer reviews of mobile application services as a case study. We propose a sentiment analysis-based framework in the first phase of our innovative two-phase framework to measure the customer satisfaction level of different generations of Apple iPhones. In the second phase, we will generate pricing strategies for the company using the outputs from the first step.

Aspect-level or aspect-based sentiment analysis (ABSA) lets companies conduct a comprehensive analysis of their customer's feedback data, enabling them to gain a deeper understanding of their customers and develop products and services that better meet their needs. ABSA is a method for categorizing data by aspect and identifying the sentiment associated with each. Aspects are the attributes or components of a product or service2. Since deep learning approaches have emerged as potential methods for achieving objectives in ABSA and their ability to capture both syntactic and semantic features of text without the need for high-level feature engineering, a comparative review of deep learning for ABSA has been provided [43]. The concerns and challenges associated with extracting distinct sentiments from various aspects and establishing relational mappings between aspects, dependencies, and interactions were highlighted in a survey. A comprehensive summary of recent developments has been provided along with their performance outcomes, demonstrating the quantitative assessment of the proposed methodology [44]. In our case study, aspects are some iPhones' features, such as the camera, operating system speed, and stylishness. Features were categorized into three groups. Software-related aspects, hardware-related aspects, and those that do not fit into one of the two groups. In other words, the aspects that can be categorized into both groups.

III. METHODOLOGY

We present a two-phase methodology for determining the lifetime profitability of each generation in a multiple-generation product line in this study. The proposed model's purpose is to generate pricing strategies that can give the company a good prediction of the actual prices and maximize overall profit over the life of MGPL. Figure 2 depicts the objectives of each step.

In the second phase of our framework, we used available partial sales data to get the sales of generations during time periods. According to Lin and Okudan [3,6], when a new generation enters the market, we allocate 60% of the sales for that quarter to the new generation and the remainder to previous generations. Prior generation sales are declining at a rate of 20% across subsequent quarters. Table I displays sales data3 for all generations within specified time periods.

TABLE I
SALES PER PRODUCT GENERATION FOR FIVE GENERATIONS OF IPHONES

Period	**Quarter**	**Gen. 8** (8/ 8 Plus)	**Gen. 9** (XS, XS Max)	**Gen. 10** (XR)	**Gen. 11** (11/ 11 Pro/ 11 ProMax)	**Gen. 12** (12/ 12 Pro/ 12 ProMax)	**Gen. 13** (13/ 13 Pro/ 13 ProMax)
1	**2019 Q1**	10,944,000	16,416,000	41,040,000			
2	**2019 Q2**	1,427,200	13,132,800	21,840,000			
3	**2019 Q3**	3,013,760	10,506,240	20,280,000			
4	**2019 Q4**		2,416,000	16,224,000	27,960,000		
5	**2020 Q1**		11,808,000	17,712,000	44,280,000		
6	**2020 Q2**			14,680,000	22,020,000		
7	**2020 Q3**			15,040,000	22,560,000		
8	**2020 Q4**				16,640,000	24,960,000	
9	**2021 Q1**				36,040,000	54,060,000	
10	**2021 Q2**				22,080,000	33,120,000	
11	**2021 Q3**				17,680,000	26,520,000	
12	**2021 Q4**					20,160,000	30,240,000

---

[2] https://monkeylearn.com/blog/aspect-based-sentiment-analysis/

[3] https://en.wikipedia.org/wiki/IPhone

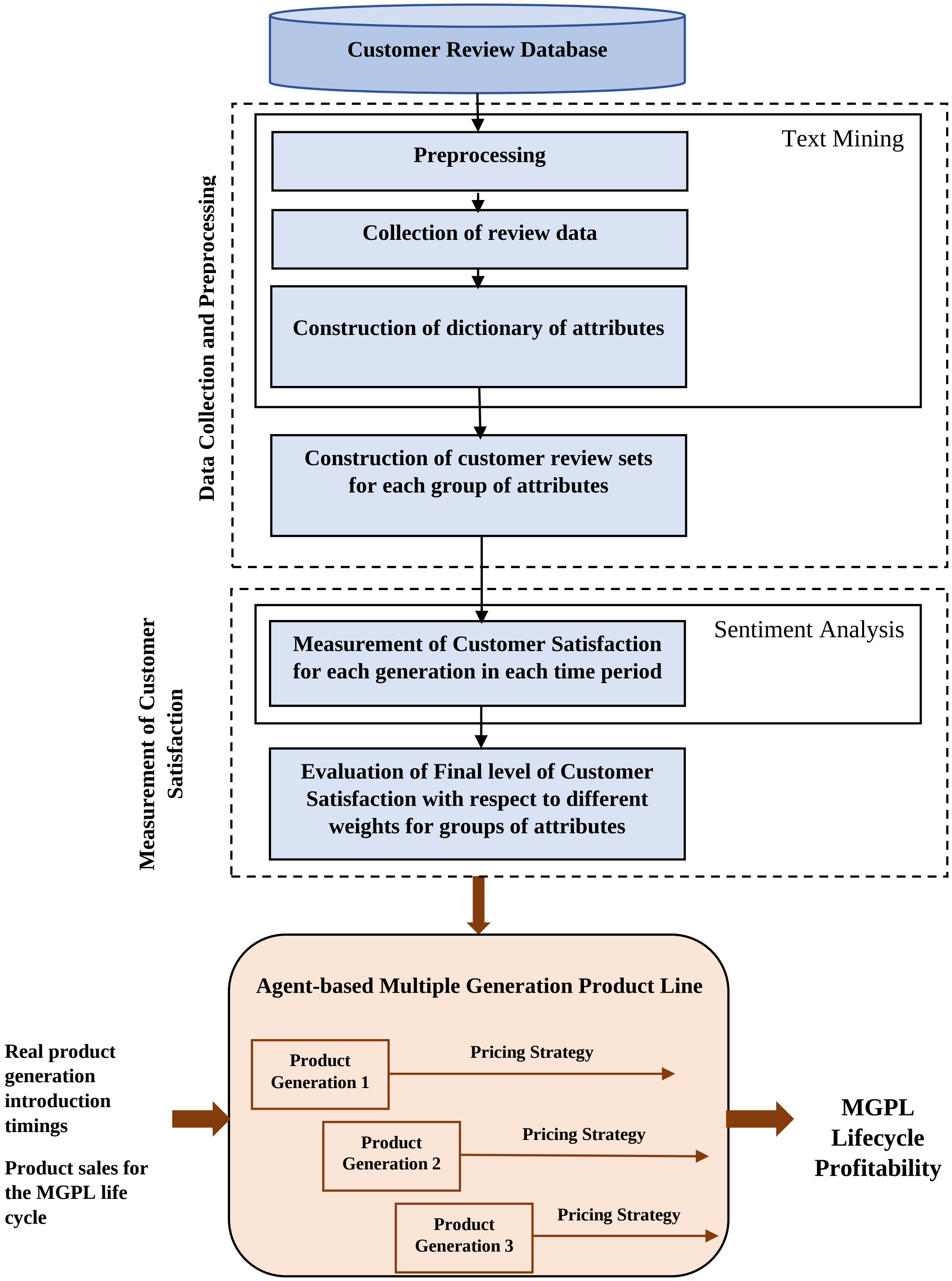


FIGURE 2. FLOW OF THE PROPOSED TWO-PHASE FRAMEWORK THAT COMBINES TEXT MINING AND SENTIMENT ANALYSIS TO PROVIDE INPUT TO AN AGENT-BASED MODEL OF A MULTIPLE GENERATION PRODUCT LINE, RESULTING IN ANALYSIS AND INSIGHTS INTO ITS LIFECYCLE PROFITABILITY

Sentiment analysis is a popular text-mining technique that reads emotional content from massive amounts of data and translates it into relevant consumer feedback [41]. Sentiment analysis has grown in popularity due to its easy-to-use results, and it has been used in various industries where user opinion is crucial [32,45]. Web-based information gives the same amount of information as sales statistics, surveys, and focus groups; however, web-based information is less structured [46]. Herein, we use sentiment analysis algorithms to analyze customer feedback from each iPhone generation and apply them to the Apple Community Forum for iPhone. The differences between iPhone versions are then investigated to trace the advancement of iPhone devices. These attempts are intended to provide an answer to the following research question: Can we improve pricing strategies by using novel sentiment analysis algorithms? Can we assist businesses in determining which aspects of their business need to be improved in the eyes of their customers?

We map sentiment-loaded words to the typical sentiment scale. The results will show the sentiment analysis's preliminary efficacy for application in multi-generation product lines. In the first stage, we use the R Programming Language to do sentiment analysis on customer reviews. The Apple community forum[4] is where consumer reviews in 2021 are gathered. The iPhone reviews are sorted and categorized into several groups based on their time and generation. To find the sentiments associated with reviews, sentiment analysis techniques are applied at the word level. When the reviews are split down into words, stop words are removed. Each review was converted into a vector of words and sentiments. As a result, at this point, a technique for assessing feelings is required.

There are numerous approaches for assessing emotion in text data. The R programming environment's "tidytext" package provides access to numerous sentiment lexicons. AFINN[5], bing[6], and nrc[7] are three general-purpose lexicons. All these lexicons are based on single words, and words are scored for positive/negative sentiment as well as emotions such as joy, anger, sadness, and so on. The "nrc" lexicon divides words into binary ("yes"/"no") categories: positive, negative, anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. The "bing" lexicon classifies terms into positive and negative categories on a binary basis. The "AFINN" lexicon gives a score to each term ranging from -5 to 5, with negative scores representing negative sentiments and positive ones representing positive sentiments[8]. Our sentiment analysis methodology takes advantage of the "bing" lexicon results. This lexicon tells us how many positive and negative terms are linked with each group.

The second phase involves the implementation of an agent-based model (ABM) with numerous agents representing different generations of the MGPL. In ABM, each generation of the MGPL is regarded as an independent agent with the ability to set its own sales price at any time. The following assumptions explain this phase: (1) When the product first hits the market, each generation has its initial price. (2) Lowering the sales price has a beneficial impact on product sales. (3) The price change's effect on product sales is believed to be known. Table II contains all the parameters and variables used in this phase. Each product generation in this model follows predetermined stochastic principles. Each agent monitors if product sales are increasing or decreasing at each time t and whether lowering the sales price would result in more profit. If the response is yes, the agent will decide to lower its price. When the product's sales price falls, the product's sales volume adjusts at a randomly chosen rate. The products of competitors are not included in this model.

TABLE II
PARAMETERS USED IN THE PROPOSED MODEL

Symbol	Meaning
$p_n^{(t)}$	Product unit price of product generation n
a_{inc}	Product sales adjustment rate when product sales are in an increasing manner
a_{dec}	Product sales adjustment rate when product sales are in a decreasing manner
C	Product unit cost
$S_n^{(t)}$	Product sales forecast of product generation n at time t
n	Product generation number, n=1, 2, …, N
d_g	Product price discount rate for general case

---

[4] https://discussions.apple.com/community/iphone/iphone_hardware?page=1

[5] From Finn Arup Nielsen

[6] From Bing Liu and collaborators

[7] From Saif Mohammad and Peter Turney

[8] https://www.tidytextmining.com/sentiment.html

d_n	Product price discount rate when a new generation of product is introduced to the market
$Pr_n^{(t)}$	Expected profit for product generation n at time t
θ	Cannibalization threshold for the product price difference, θ<1
$Can(t)$	Cannibalization sales reduction rate
$TP_m^{(t)}$	M different conditions of total profit for product generation (n-1) and n at time period t
$TM_n^{(t)}$	Customer satisfaction parameter for generation n at time t
$f1_n^{(t)}$	Number of Positive terms in reviews associated with generation n at time t
$f2_n^{(t)}$	Number of Negative terms in reviews associated with generation n at time t

If a new generation joins the market, all prior generations, except for the last, will reduce the price of their items at a predetermined rate. Furthermore, the overall volume of all product sales should not exceed the company's production capacity within any given time. Finally, we anticipate gaining some useful pricing strategies for each generation of the multiple-generation product line at each stage of its existence. We anticipate that in comparison to Lin et al.'s work [7], we will find superior Pricing Strategies for each generation of the MGPL. We expect higher sales prices for generations with more positive reviews and lower prices for generations with fewer positive reviews.

We considered $f1_n^{(t)}$ and $f2_n^{(t)}$as the number of positive and negative terms that come from using the Bing lexicon on the related reviews. Related reviews include generation *n* of iPhone and submitted at time *t*. $TM_n^{(t)}$ will be as follows:

$$TM_n^{(t)} = a_1 * f1_n^{(t)} + a_2 * f2_n^{(t)}. \tag{1}$$

A simple linear model was used to create the $TM_n^{(t)}$ formula. This method employs direct counts of comments rather than the ratio of positive versus negative comments or the ratio of positive comments overall. a_1 and a_2 are coefficients (weights) of $f1_n^{(t)}$ and $f2_n^{(t)}$ in the linear model. Different values for two parameters in the linear model will be tested and evaluated by RMSE to find the best values. As a result, the more comments about a product, the greater the upward adjustment to the price predicted by ABM. Moreover, the number of negative reviews can increase the price on their own. We must emphasize that as more data becomes available, these weights will not only need to be adjusted, but it also implies the need for a more complicated formula in the future to avoid above mentioned problems. For example, one generalization could be:

$$TM_n^{(t)} = a_1 * \min\left\{B_1, f1_n^{(t)}\right\} + a_2 * \min\left\{B_2, f2_n^{(t)}\right\}. \tag{2}$$

Negative comments could never increase the price on their own if $B_2 < 1/a_2$. This above formula necessitates fitting values for four parameters. We leave such a more complex formula as future work due to the lack of enough data at the present time. We also emphasize that the main point of this work is to determine whether sentiment analysis explains most of the differences between the ABM and the real prices. With additional data, it may be possible to improve this formula.

Decision rules for previous generations (k-2, k-3, …, 1)

In time period t, with the latest generation k:

Step 1: sales are increasing.

$$S_n^{(t)} - S_n^{(t-1)} \geq 0$$

For n< k-1 (k-2, k-3, …, 1) if:

$$\left[\boldsymbol{P}_{\boldsymbol{n}}^{(\boldsymbol{t-1})} - \boldsymbol{C}\right] \times \boldsymbol{S}_{\boldsymbol{n}}^{(\boldsymbol{t})} \leq \{\left[\boldsymbol{P}_{\boldsymbol{n}}^{(\boldsymbol{t-1})} \times \boldsymbol{d}_{\boldsymbol{g}}\right] - \boldsymbol{C}\} \times \boldsymbol{S}_{\boldsymbol{n}}^{(\boldsymbol{t})} \times \boldsymbol{a}_{\boldsymbol{inc}}$$

And $\boldsymbol{P}_{\boldsymbol{n}}^{(\boldsymbol{t-1})} \times \boldsymbol{d}_{\boldsymbol{g}} > \boldsymbol{C}$

Then,

$$\boldsymbol{P_n^{(t)} = P_n^{(t-1)} \times d_g \times TM_n^{(t-1)}} \tag{3}$$

Otherwise,

$$\boldsymbol{P_n^{(t)} = P_n^{(t-1)} \times TM_n^{(t-1)}} \tag{4}$$

Step 2: sales are decreasing.

$$S_n^{(t)} - S_n^{(t-1)} < 0$$

For n< k-1 (k-2, k-3, …, 1) if:

$$\boldsymbol{\left[P_n^{(t-1)} - C\right] \times S_n^{(t)} \leq \{\left[P_n^{(t-1)} \times d_g\right] - C\} \times S_n^{(t)} \times a_{dec}}$$

And $\boldsymbol{P_n^{(t-1)} \times d_g > C}$
Then,

$$\boldsymbol{P_n^{(t)} = P_n^{(t-1)} \times d_g \times TM_n^{(t-1)}} \tag{5}$$

Otherwise,

$$\boldsymbol{P_n^{(t)} = P_n^{(t-1)} \times TM_n^{(t-1)}} \tag{6}$$

Step 3: When a new generation k+1 comes to the market, older ones n< k+1 adjust their prices:

$$P_n^{(t)} = P_n^{(t)} \times d_n \tag{7}$$

For two recent generations (k and k-1)

We assume that cannibalization has just occurred for the last two generations. Three new parameters have been added to the model. The cannibalization threshold for the product price difference ϴ. The cannibalization sales reduction rate $Can(t)$ and $TP_m(t)$, where m denotes various conditions of total profits for the last two product generations at time t. In other words, m states are indicative of market circumstances. There are four different cases:

1) Discount on previous generation (k-1)
2) Discount on current generation (k)
3) Discount on both generations (k-1, k)
4) No discount on both generations.

The case with the maximum profit will be selected and continued. In each case, we should check whether the sales price ratio between the previous generation and the current one is above ϴ (cannibalization threshold).

Step 1: the sales are in an increasing manner.

$$S_n^{(t)} - S_n^{(t-1)} \geq 0$$

$$\text{For } n \geq k-1 \ (k-1, k)$$

Case 1: No price or sales volume changes across two generations, and cannibalization may occur based on ϴ.
If:

$$[P_{k-1}^{(t-1)} < P_k^{(t-1)} \ and \ (\frac{P_{k-1}^{(t-1)}}{P_k^{(t-1)}}) < \theta]$$

Or,

$$[P_{k-1}^{(t-1)} > P_k^{(t-1)} \ and \ (\frac{P_k^{(t-1)}}{P_{k-1}^{(t-1)}}) > \theta]$$

$$TP_1(t) = \left[P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} + [P_k^{(t-1)} \times TM_k^{(t-1)} - C] \times S_k^t,$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times TM_k^{(t-1)} \tag{8}$$

Otherwise,

$$(\frac{P_{(k-1)}^{(t-1)}}{P_k^{(t-1)}}) < \theta$$

Then:

$$TP_1(t) = \left[P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} \times [1 + can(t)] + \left[P_k^{(t-1)} \times TM_k^{(t-1)} - C\right] \times S_k^{(t)} \times [1 - can(t)],$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times TM_k^{(t-1)} \tag{9}$$

Or:

$$(\frac{P_k^{(t-1)}}{P_{k-1}^{(t-1)}}) < \theta$$

$$TP_1(t) = \left[P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} \times [1 - can(t)] + \left[P_k^{(t-1)} \times TM_k^{(t-1)} - C\right] \times S_k^{(t)} \times [1 + can(t)]$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times TM_k^{(t-1)} \tag{10}$$

Case 2: Discount on the previous generation (k-1) with the discount rate d_g. (The previous generation may be so much cheaper as customers prefer the previous generation to the new one, so the new generation will be cannibalized by the previous one.)

$$(\frac{[P_{k-1}^{(t-1)} \times d_g]}{P_k^{(t-1)}}) > \theta$$

Then:

$$TP_2(t) = \left[P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} \times a_{inc} + \left[P_k^{(t-1)} \times TM_k^{(t-1)} - C\right] \times S_k^{(t)},$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times TM_k^{(t-1)} \tag{11}$$

Otherwise,

$$TP_2(t) = \left[P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} \times a_{inc} \times [1 + can(t)] + \left[P_k^{(t-1)} \times TM_k^{(t-1)} - C\right] \times S_k^{(t)} \times [1 - can(t)]$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times TM_k^{(t-1)} \tag{12}$$

Case 3: Discount on the new generation (k) with the discount rate d_g. (The new generation may be so much cheaper as customers prefer the new generation to the previous one, so the previous generation will be cannibalized by the new one.)

If:

$$\left(\frac{P_{k-1}^{(t-1)}}{[P_k^{(t-1)} \times d_g]}\right) < [\frac{1}{\theta}]$$

Then:

$$TP_3(t) = \left[P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} + \left[P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} - C\right] \times S_k^{(t)} \times a_{inc}$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} \quad (13)$$

Otherwise,

$$TP_3(t) = \left[P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} \times [1 - can(t)] + \left[P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} - C\right] \times S_k^{(t)} \times a_{inc} \times [1 + can(t)]$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} \quad (14)$$

Case 4: In this case, there is a discount on both recent generations (k, k-1), and they will both enter cannibalization circumstances.

If:

$$\left(\frac{P_{k-1}^{(t-1)} \times d_g}{[P_k^{(t-1)} \times d_g]}\right) > \theta \text{ and } \left(\frac{P_k^{(t-1)} \times d_g}{[P_{k-1}^{(t-1)} \times d_g]}\right) > \theta$$

Then:

$$TP_4(t) = \left[P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} \times a_{inc} + \left[P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} - C\right] \times S_k^{(t)} \times a_{inc}$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} \quad (15)$$

Otherwise,

$$\left(\frac{P_{k-1}^{(t-1)} \times d_g}{[P_k^{(t-1)} \times d_g]}\right) < \theta$$

$$TP_4(t) = \left[P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} \times a_{inc} \times [1 + can(t)] + \left[P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} - C\right] \times S_k^{(t)} \times a_{inc} \times [1 - can(t)]$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} \quad (16)$$

Or:

$$\left(\frac{P_k^{(t-1)} \times d_g}{[P_{k-1}^{(t-1)} \times d_g]}\right) < \theta$$

Then:

$$TP_4(t) = \left[P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} \times a_{inc} \times [1 - can(t)] + \left[P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} - C\right] \times S_k^{(t)} \times a_{inc} \times [1 + can(t)]$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} \quad (17)$$

The case with Max $TP_m(t)$, m= 1 to 4 will be chosen.

Step 2: the sales are in a decreasing manner.

$$S_n^{(t)} - S_n^{(t-1)} < 0$$

$$\text{For } n \geq k - 1 \ (k - 1, k)$$

Case 1: No price or sales volume changes across two generations, and cannibalization may occur based on ϴ.
If:

$$[P_{k-1}^{(t-1)}/P_k^{(t-1)}] > \theta$$

Then:

$$TP_1(t) = \left[P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} + [P_k^{(t-1)} \times TM_k^{(t-1)} - C] \times S_k^{t},$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times TM_k^{(t-1)} \tag{18}$$

Otherwise,

$$(\frac{P_{(k-1)}^{(t-1)}}{P_k^{(t-1)}}) < \theta$$

Then:

$$TP_1(t) = \left[P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} \times [1 + can(t)] + \left[P_k^{(t-1)} \times TM_k^{(t-1)} - C\right] \times S_k^{(t)} \times [1 - can(t)],$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times TM_k^{(t-1)} \tag{19}$$

Or:

$$(\frac{P_k^{(t-1)}}{P_{k-1}^{(t-1)}}) < \theta$$

$$TP_1(t) = \left[P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} \times [1 - can(t)] + \left[P_k^{(t-1)} \times TM_k^{(t-1)} - C\right] \times S_k^{(t)} \times [1 + can(t)]$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times TM_k^{(t-1)} \tag{20}$$

Case 2: Discount on the previous generation (k-1) with the discount rate d_g. (The previous generation may be so much cheaper as customers prefer the previous generation to the new one, so the new generation will be cannibalized by the previous one.)

$$(\frac{[P_{k-1}^{(t-1)} \times d_g]}{P_k^{(t-1)}}) > \theta$$

Then:

$$TP_2(t) = \left[P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} \times a_{dec} + \left[P_k^{(t-1)} \times TM_k^{(t-1)} - C\right] \times S_k^{(t)},$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times TM_k^{(t-1)} \tag{21}$$

Otherwise,

$$TP_2(t) = \left[P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} \times a_{dec} \times [1 + can(t)] + \left[P_k^{(t-1)} \times TM_k^{(t-1)} - C\right] \times S_k^{(t)} \times [1 - can(t)]$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times TM_k^{(t-1)} \tag{22}$$

Case 3: Discount on the new generation (k) with the discount rate d_g. (The new generation may be so much cheaper as customers prefer the new generation to the previous one, so the previous generation will be cannibalized by the new one.)

If:

$$\left(\frac{P_{k-1}^{(t-1)}}{[P_k^{(t-1)} \times d_g]}\right) < [\frac{1}{\theta}]$$

Then:

$$TP_3(t) = \left[P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} + \left[P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} - C\right] \times S_k^{(t)} \times a_{dec}$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} \tag{23}$$

Otherwise,

$$TP_3(t) = \left[P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} \times [1 - can(t)] + \left[P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} - C\right] \times S_k^{(t)} \times a_{dec} \times [1 + can(t)]$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} \tag{24}$$

Case 4: In this case, there is a discount on both recent generations (k, k-1), and they will both enter cannibalization circumstances.

If:

$$\left(\frac{P_{k-1}^{(t-1)} \times d_g}{[P_k^{(t-1)} \times d_g]}\right) > \theta \text{ and } \left(\frac{P_k^{(t-1)} \times d_g}{[P_{k-1}^{(t-1)} \times d_g]}\right) > \theta$$

Then:

$$TP_4(t) = \left[P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} \times a_{dec} + \left[P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} - C\right] \times S_k^{(t)} \times a_{dec}$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} \tag{25}$$

Otherwise,

$$\left(\frac{P_{k-1}^{(t-1)} \times d_g}{[P_k^{(t-1)} \times d_g]}\right) < \theta$$

$$TP_4(t) = \left[P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} \times a_{dec} \times [1 + can(t)] + \left[P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} - C\right] \times S_k^{(t)} \times a_{dec} \times [1 - can(t)]$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} \tag{26}$$

Or:

$$\left(\frac{P_k^{(t-1)} \times d_g}{[P_{k-1}^{(t-1)} \times d_g]}\right) < \theta$$

Then:

$$TP_4(t) = \left[P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} \times a_{dec} \times [1 - can(t)] + \left[P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} - C\right] \times S_k^{(t)} \times a_{dec} \times [1 + can(t)]$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} \tag{27}$$

The case with Max $TP_m(t)$, *m*= 1 to 4 will be chosen.

Step 3: When a new product generation (*n*= *k*+1) comes to the market, the previous ones (*n*< *k*+1) adjust their unit sales price to be:

$$p_n^{(t)} = P_n^{(t)} \times d_n \tag{28}$$

IV. RESULTS

A. Text Mining Result

In the first phase we extract 1591 reviews for 2021 from the Apple community forum. The reviews were first filtered based on containing the word iPhone, as we do not want to include reviews associated with other products such as MacBooks and iPods. The reviews were then split based on their time into four groups according to yearly quarters. Each of the four groups is further split into different subgroups based on the product generations. We thus have different groups of reviews with time and generation. Reviews are preprocessed and tokenized. Using the Bing lexicon for the reviews in each group, we can get the number $f1_n^{(t)}$ of positive and $f2_n^{(t)}$ of negative terms. Recall that we define a sentiment analysis factor $TM_n^{(t)}$ as a linear combination of these counts: $TM_n^{(t)} = a_1 * f1_n^{(t)} + a_2 * f2_n^{(t)}$. Here a_1 and a_2 are weights determined by minimizing the RMSE of the difference between model predictions of price and the observed price. In this case, the best weights were obtained as $a_1 = 0.1269$ and $a_2 = 0.0340$, resulting in the following linear expression for $TM_n^{(t)}$:

$$TM_n^{(t)} = 0.1269 * f1_n^{(t)} + 0.0340 * f2_n^{(t)}$$

Table III reports the TM parameter value for each time and generation. Note that when there are no reviews, the parameter is automatically set equal to one, as no adjustments should be made to the price according to reviews if no reviews are available.

TABLE III
NUMBER OF POSITIVE AND NEGATIVE TERMS AND THE SENTIMENT ANALYSIS PARAMETERS FOR DIFFERENT GENERATIONS IN DIFFERENT TIMES

	2021-Q1		**2021-Q2**		**2021-Q3**		**2021-Q4**	
	Gen 11	**Gen 12**	**Gen 11**	**Gen 12**	**Gen 11**	**Gen 12**	**Gen 12**	**Gen 13**
$f1_n^{(t)}$	-	-	3	-	3	8	58	54
$f2_n^{(t)}$	-	-	21	-	12	3	95	80
$TM_n^{(t)}$	1	1	1.09	1	1.16	1.11	10.59	9.57

Table IV illustrates the sensitivity of the prediction RMSE to the different values of weights. We note that how the text mining results are used greatly influences the quality of the predictions. By tuning the weights, the RMSE can be reduced by two orders of magnitude.

TABLE IV
RMSE FOR DIFFERENT VALUES OF THE WEIGHTS a_1, a_2 USED TO CALCULATE THE SENTIMENT ANALYSIS PARAMETER

RMSE		$a_2 =$				
		0.014	**0.024**	**0.034**	**0.044**	**0.054**
$a_1 =$	**0.0269**	504.8	441.7	389.9	354.6	312.2
	0.0769	330.2	290.6	258.8	236.8	191.3
	0.1269	191.4	99.9	9.1	24.4	32.1
	0.1769	193.8	165.8	138.7	172.6	204.1

Further analysis of the reviews provides additional insights. Topic modeling was done on the reviews with the different number of topics, and five is the number of topics that fits the data best. We experimented with the number of topics, and five was chosen because the frequent terms in each topic were mostly related, and the result was more meaningful. From the most frequent terms in each topic, we removed terms such as verbs and adverbs and concentrated on terms that refer to features of iPhones. Figure 3 shows the most frequent terms in each topic.

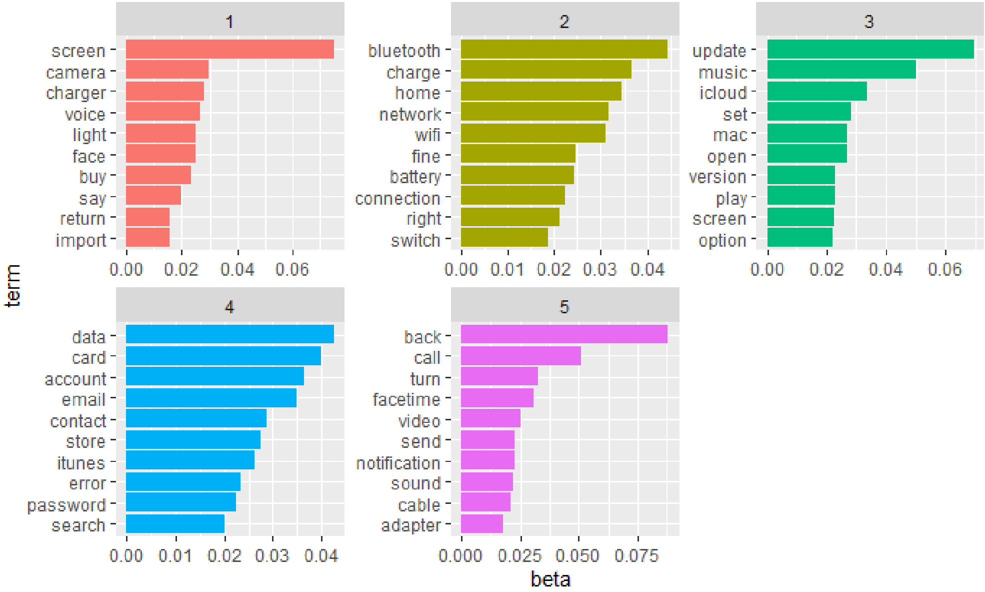


FIGURE 3. TOPIC MODELLING OUTPUT WITH 5 TOPICS

These most frequent terms (features) have been categorized into three groups: (1) hardware-related features, (2) software-related features, and (3) between-group features. The first and second group contains features related to iPhone hardware and software respectively. The terms in the third group are those that refer to both the hardware and software of iPhones, like Bluetooth, which can declare something about hardware or software. The reviews are split into these three groups and further divided based on the product generations. This results in the number of positive and negative terms in each group for different generations. Figures 4-6 show the details from our analysis for generations in each of the three groups.

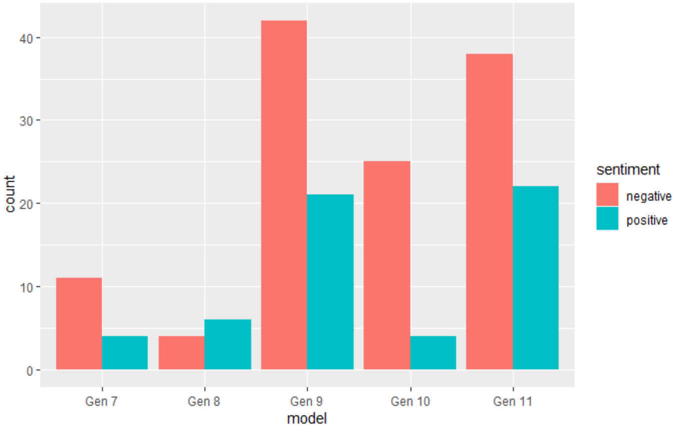


FIGURE 4. NUMBER OF POSITIVE AND NEGATIVE TERMS FOR DIFFERENT GENERATION EXTRACTED FROM REVIEWS ASSOCIATED WITH HARDWARE FEATURES

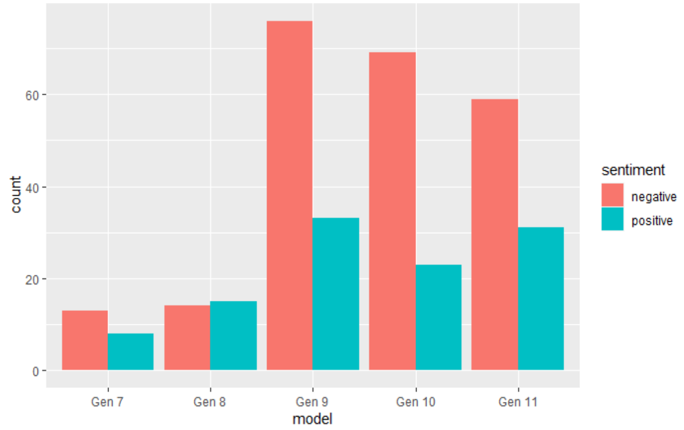


FIGURE 5. NUMBER OF POSITIVE AND NEGATIVE TERMS FOR DIFFERENT GENERATION EXTRACTED FROM REVIEWS ASSOCIATED WITH SOFTWARE FEATURES

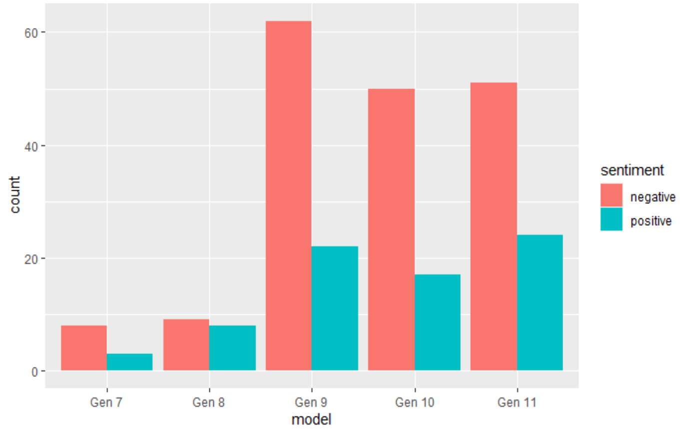


FIGURE 6. NUMBER OF POSITIVE AND NEGATIVE TERMS FOR DIFFERENT GENERATION EXTRACTED FROM REVIEWS ASSOCIATED WITH BETWEEN-GROUP FEATURES

After this step, with the help of Bing lexicon results, we can get three *F* Parameters for hardware, software, and between-group of each generation. For each of these three groups, the *F* parameter is the ratio of positive reviews to all reviews in that relative group:

$$F_{(n)} = \frac{\sum_t f1_n^{(t)}}{\sum_t f1_n^{(t)} + \sum_t f2_n^{(t)}}. \tag{29}$$

Table V shows the values of *F* parameters for different generations in 2021. We also suggest a total *F* parameter, which will be calculated below. Since we do not have a basis for giving different weights to different components, we use an unweighted average:

$$F_{total(n)} = \frac{1}{3} \times F_{Hardware(n)} + \frac{1}{3} \times F_{Software(n)} + \frac{1}{3} \times F_{Between-group(n)}. \tag{30}$$

TABLE V
ALL F PARAMETERS FOR DIFFERENT GENERATIONS IN 2021

	$\boldsymbol{F_{Hardware(n)}}$	$\boldsymbol{F_{Software(n)}}$	$\boldsymbol{F_{Between-group(n)}}$	$F_{total(n)}$
Gen 7	0.27	0.38	0.27	0.307
Gen 8	0.6	0.52	0.47	0.53
Gen 9	0.33	0.3	0.26	0.297
Gen 10	0.14	0.25	0.25	0.213
Gen 11	0.37	0.34	0.32	0.343

These results show the company how much consumers in 2021 were satisfied with each iPhone generation. The values reported in Table V show that generation 8 has the highest F_{total} parameter. So, this generation has the highest level of customer satisfaction in comparison with other generations. After that, generations 11, 7, 9, and 10 have the next levels of customer satisfaction, respectively. It can help companies find which generations need more improvement or how much customers are satisfied with each generation of iPhones. Moreover, F parameters related to hardware, software, and between-group features show the company which of these three different areas needs more improvement. For example, in generation 10 parameters, $F_{Hardware}$ is the lowest one in comparison with the others. It can help companies to find that for generation 10; they need to focus on hardware more than software and between group features depending on customer's opinions.

B. ABM Result

With all the available agent-based-model (ABM) inputs, we can get the pricing strategies for different generations in each time period. Using the proposed ABM, companies can get the best pricing strategies with the highest profits for different generations in each quarter of 2021. We used the NetLogo software to simulate the model. We used the sales data in Table I in the Agent-based model, and one sale unit equals 1,000,000 sale volume in real data. Table VI summarizes the parameters used in the simulation experiment.

TABLE VI
PARAMETERS USED IN THE SIMULATION EXPERIMENT.

Symbol	**Meaning**	**Simulation Value**
$p_n^{(t)}$	Product unit price of product generation n	-
a_{inc}	Product sales adjustment rate when product sales are in an increasing manner	0.5
a_{dec}	Product sales adjustment rate when product sales are in a decreasing manner	0.7
C	Product unit cost	-
d_g	Product price discount rate for general case	0.7
d_n	Product price discount rate when a new generation of product is introduced to the market	0.8
θ	Cannibalization threshold for the product price difference, θ<1	0.4
$Can(t)$	Cannibalization sales reduction rate	0.6

Real initial prices were used for $p_n^{(t)}$ when a new generation came to the market for the first time. The costs of different generations were assumed to be known based on a linear relation with the initial prices. Different values for d_g and d_n were tested and evaluated by considering the fit between the prices predicted by the model and the observed prices. The values that minimized the RMSE were found to be 0.7 and 0.8 for d_g and d_n, respectively. In the simulation experiment, the total profit from the current and previous generations is assessed under the four cannibalization conditions. In the first experiment, we used the base ABM model with the above parameters to monitor the total profit of the two latest generations under four cannibalization scenarios. Figure 7 illustrates the total profit from the two latest generations under different scenarios.

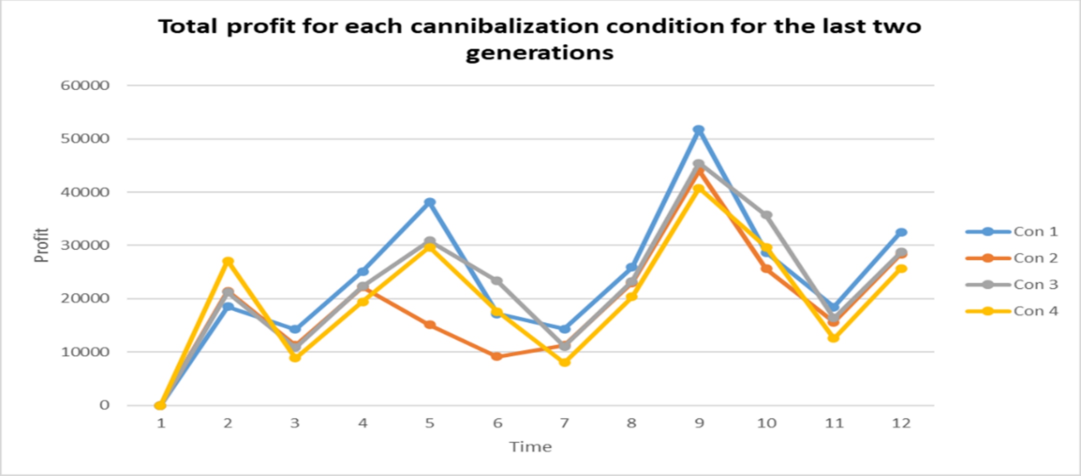


FIGURE 7. TOTAL PROFIT FOR EACH CANNIBALIZATION CONDITION FOR THE LAST TWO GENERATIONS

In simulation experiment 2, we simulated the model with the defined parameters in Table VI, and we added the TM parameter to the model with the related values from Table III. We considered $TM_n^{(t)} = 1$ whenever we did not have a determined value for the TM parameter. Figure 8 shows the results of this second experiment.

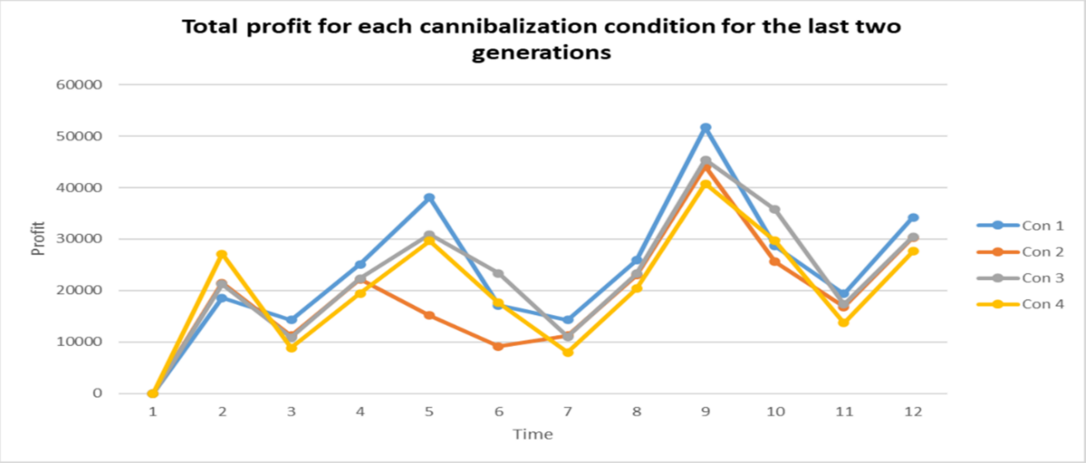


FIGURE 8. TOTAL PROFIT FOR EACH CANNIBALIZATION CONDITION FOR THE LAST TWO GENERATIONS

The results show that two cannibalization conditions (conditions 1 and 3) generate more profit in comparison with others (conditions 2 and 4). Condition 1 is when there is no discount on both generations, but cannibalization occurs based on the cannibalization threshold. Condition 3 is the case where there is a discount on the most recent generation, which results in the new generation cannibalizing the previous one.

After adding the defined parameters to the model, results show that at the latest periods, with the help of the text mining parameter, the profit was increased. Moreover, the simulation gives us price changes based on choosing the best strategy with the maximum profit in each time period. In the table below, the price changes for generations have been listed from the first simulation experiments (without using the text mining parameter) and compared with the actual prices over time.

TABLE VII
PRICE CHANGES DURING TIME FROM SIMULATION EXPERIMENTS IN COMPARISON WITH REAL ONES

Price changes													
Time		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Gen 8	Predicted price without TM	650	650	650									
	Actual Price	650	650	650									
Gen 9	Predicted price without TM	1049	1049	1049									
	Actual Price	1049	1049	1049									
Gen 10	Predicted price without TM	749	749	749	599.2	599.2	599.2	599.2					
	Actual Price	749	749	749	600	600	600	600					
Gen 11	Predicted price without TM				932	932	932	932	745.6	745.6	745.6	745.6	
	Actual Price				932	932	932	932	830	830	830	830	
Gen 12	Predicted price without TM								976	976	976	976	780.8
	Actual Price								976	976	976	976	880
Gen 13	Predicted price without TM												976
	Actual Price												976

Regarding the availability of text mining parameters for some generations in the latest periods, we can see the changes for Gen 5 and 6 in Table VIII. The actual prices were extracted from the GSMArena[9] website.

TABLE VIII
PRICE CHANGES DURING THE LATEST TIME PERIODS FROM TWO SIMULATION EXPERIMENTS IN COMPARISON WITH REAL ONES

Price changes			
Time		**11**	**12**
	Predicted price without TM	**745.6**	
Gen 11	Predicted price with TM	**816.2**	
	Actual price	**830**	
Gen 12	Predicted price without TM	**976**	**780.8**
	Predicted price with TM	**976**	**872.3**
	Actual price	**976**	**880**
Gen 13	Predicted price without TM		**976**
	Predicted price with TM		**976**
	Actual price		**976**

Table VIII shows that the text mining parameter explains most of the difference between the ABM prediction and the actual prices. These modifications are demonstrated by the reduction in RMSE from 75.19 to 9.13 or an almost 88% reduction in RMSE. Furthermore, as we carefully tuned the key parameters to reduce the RMSE of the ABM model on its own, we claim that this type of decrease could not be accomplished merely by better implementing ABM without the TM and that combining the ABM with text mining resulted in the improvement.

---

[9] https://www.gsmarena.com/charts_show_the_evolution_of_iphone_prices_over_the_years_and_show_how_well_older_phones_held_their_-news-51097.php#image0

V. DISCUSSION

This paper builds on the two-phased method proposed by Kilicay-Ergin, Lin, and Okudan's study, where the authors developed pricing strategies for MGPs to maximize their lifetime profitability using their proposed two-phase methodology [8]. Although they have good pricing strategies, there are several circumstances that can compel corporations to adjust the price of their products, affecting the company's profit. Customer satisfaction is one of them. This additional parameter was employed in our suggested ABM to simulate cannibalization conditions in a more realistic manner. The outcome indicates that the price will be changed by this new parameter. The customer opinion toward each generation of an MGPL in each time period might affect the price of that generation in the next time period, as well as the profit. In comparison to [8], we have more data input for our suggested ABM, and adding a new effective parameter will provide us with more realistic pricing strategies. As a result, online evaluations and their sentiments can assist businesses in gaining a better understanding of their customers and how they interpret successive generations of the MGPL.

VI. CONCLUSION AND FUTURE WORK

The MGPL planning problem is a multidimensional problem with several dimensions. Behavioral models are generally concerned with sales forecasts and generational introduction timing, whereas dynamic competition models are concerned with pricing dynamics under competition to determine optimal pricing options for the most recent generations. To investigate the entire MGPL planning challenge, we present a theoretical two-phase approach. We assess relevant reviews to determine customer satisfaction and ultimately obtain dynamic pricing strategies for the full multiple-generation product line. The model views a multiple-generation product line as an adaptive living creature whose behavior is affected by the strategic decisions made by each product generation. In this research, an agent-based model of MPGL pricing decisions is created to assess the dynamics of alternative pricing strategies on the overall profitability of the MGPL. Experiments in simulation provide insights into the dynamics of pricing strategies under various cannibalization scenarios. For pricing decisions, four alternative cannibalization scenarios are investigated in this study. Analyses of the results show that higher customer satisfaction leads to higher prices and more profit, whereas lower customer satisfaction forces companies to drop prices and improve the product.

It is vital to highlight that the results are valid in circumstances when the company operates in a monopoly market and new generations are introduced to answer varied customer value preferences and expectations. Early in the MGPL lifecycle, customer product selection is based on product performance evaluation. Now, customer valuation moves to product quality-price preferences. As a result, profits under cannibalization should be considered in the later lifecycle states of the MGPL. In this work, we used partial sales data and real introduction timings to get the pricing strategies for different time periods. In future work, we will use the DSVM model from [8] to predict sales and introduction timings of generations in the whole MGPL life cycle.

If there are several competitors in the market, the ABM outcomes may alter since sales dynamics would change because of competition. Future research will investigate the dynamics of four cannibalization scenarios under a competitive market environment. The current model is the initial step in analyzing the pricing dynamics of MGPL. There are other factors that influence an MPGL's sales volume and cannibalization circumstances. For instance, the products launched by competitors, market conditions such as technological developments, and product quality all influence the demand dynamics of each generation. In future works, we intend to develop this model and integrate other aspects into the study, such as many competitors and technology-changing factors. It is hoped that this form of MPGL agent-based modeling will serve as a test bed for analyzing various aspects of the complex cannibalization dynamics of an MPGL planning problem.

[1] Krankel, R. M., Duenyas, I., & Kapuscinski, R. (2006). Timing successive product introductions with demand diffusion and stochastic technology improvement. Manufacturing & Service Operations Management, 8(2), 119-135.

[2] Morgan, L. O., Morgan, R. M., & Moore, W. L. (2001). Quality and time-to-market trade-offs when there are multiple product generations. Manufacturing & Service Operations Management, 3(2), 89-104.

[3] Lin, C.-Y., & Kremer, G. E. O. (2014). Strategic decision making for multiple-generation product lines using dynamic state variable models: The cannibalization case. Computers in Industry, 65(1), 79-90.

[4] Jiao, J., & Tseng, M. M. (1999). A methodology of developing product family architecture for mass customization. Journal of Intelligent Manufacturing, 10(1), 3-20.

[5] Simpson, T. W. (2004). Product platform design and customization: Status and promise. Ai Edam, 18(1), 3-20.

[6] Lin, C.-Y., & Okudan, G. E. (2013). Planning for multiple-generation product lines using dynamic variable state models with data input from similar products. Expert systems with applications, 40(6).

[7] Lin, C.-y., Kilicay-Ergin, N. H., & Okudan, G. E. (2011). Agent-based modeling of dynamic pricing scenarios to optimize multiple-generation product lines with cannibalization. Procedia Computer Science, 6, 311-316.

[8] Kilicay-Ergin, N., Lin, C.-y., & Okudan, G. E. (2015). Analysis of dynamic pricing scenarios for multiple-generation product lines. Journal of Systems Science and Systems Engineering, 24(1), 107-129.

[9] Ofek, E., & Sarvary, M. (2003). R&D, marketing, and the success of next-generation products. Marketing Science, 22(3), 355-370.

[10] Edelheit, L. S. (2004). The IRI Medalist's Address: Perspective On GE Research & Development. Research-Technology Management, 47(1), 49-55.

[11] Norton, J. A., & Bass, F. M. (1987). A diffusion theory model of adoption and substitution for successive generations of high-technology products. Management science, 33(9), 1069-1086.

[12] Mahajan, V., & Muller, E. (1996). Timing, diffusion, and substitution of successive generations of technological innovations: The IBM mainframe case. Technological Forecasting and Social Change, 51(2), 109-132.

[13] Bardhan, A., & Chanda, U. (2008). A model for first and substitution adoption of successive generations of a product. International Journal of Modelling and simulation, 28(4), 487-494.

[14] Huang, C.-Y., & Tzeng, G.-H. (2008). Multiple generation product life cycle predictions using a novel two-stage fuzzy piecewise regression analysis method. Technological forecasting and social change, 75(1), 12-31.

[15] Dobson, G., & Kalish, S. (1988). Positioning and pricing a product line. Marketing Science, 7(2), 107-125.

[16] Arslan, H., Kachani, S., & Shmatov, K. (2009). Optimal product introduction and life cycle pricing policies for multiple product generations under competition. Journal of Revenue and Pricing Management, 8(5), 438-451.

[17] Li, H., & Graves, S. C. (2012). Pricing decisions during inter-generational product transition. Production and Operations Management, 21(1), 14-28.

[18] Gallego, G., & Wang, R. (2014). Multiproduct price optimization and competition under the nested logit model with product-differentiated price sensitivities. Operations Research, 62(2), 450-461.

[19] Kim, N., Srivastava, R. K., & Han, J. K. (2001). Consumer decision-making in a multi-generational choice set context. Journal of Business Research, 53(3), 123-136.

[20] Schön, C. (2010). On the optimal product line selection problem with price discrimination. Management Science, 56(5), 896-902.

[21] Chen, J.-M., & Chang, C.-I. (2013). Dynamic pricing for new and remanufactured products in a closed-loop supply chain. International Journal of Production Economics, 146(1), 153-160.

[22] Fruchter, G., Fligler, A., & Winer, R. (2006). Optimal product line design: Genetic algorithm approach to mitigate cannibalization. Journal of optimization theory and applications, 131(2), 227-244.

[23] Hotho, A., Nürnberger, A., & Paaß, G. (2005). A brief survey of text mining. Ldv Forum,

[24] Zhan, J., Loh, H. T., & Liu, Y. (2009). Gather customer concerns from online product reviews–A text summarization approach. Expert Systems with Applications, 36(2), 2107-2115.

[25] Thorleuchter, D., Van den Poel, D., & Prinzie, A. (2010). Mining ideas from textual information. Expert Systems with Applications, 37(10), 7182-7188.

[26] Nasukawa, T., & Yi, J. (2003). Sentiment analysis: Capturing favorability using natural language processing. Proceedings of the 2nd international conference on Knowledge capture,

[27] Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. Proceedings of human language technology conference and conference on empirical methods in natural language processing,

[28] Beineke, P., Hastie, T., & Vaithyanathan, S. (2004). The sentimental factor: Improving review classification via human-provided information. Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04),

[29] Pang, B., & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. arXiv preprint cs/0409058.

[30] Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. arXiv preprint cs/0205070.

[31] Turney, P. D. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. arXiv preprint cs/0212032.

[32] Hu, M., & Liu, B. (2004a). Mining and summarizing customer reviews. Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining,

[33] Hu, M., & Liu, B. (2004b). Mining opinion features in customer reviews. AAAI,

[34] Kamps, J., & Marx, M. (2001). Words with attitude.

[35] Zhang, W., Xu, H., & Wan, W. (2012). Weakness Finder: Find product weakness from Chinese reviews by using aspects based sentiment analysis. Expert Systems with Applications, 39(11), 10283-10291.

[36] Kang, D., & Park, Y. (2012). Measuring customer satisfaction of service based on an analysis of the user generated contents: Sentiment analysis and aggregating function based MCDM approach. 2012 IEEE International Conference on Management of Innovation & Technology (ICMIT).

[37] Täckström, O., & McDonald, R. (2011). Semi-supervised latent variable models for sentence-level sentiment analysis. The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies,

[38] Mostafa, M. M. (2013). More than words: Social networks' text mining for consumer brand sentiments. Expert systems with applications, 40(10), 4241-4251.

[39] Kontopoulos, E., Berberidis, C., Dergiades, T., & Bassiliades, N. (2013). Ontology-based sentiment analysis of twitter posts. Expert systems with applications, 40(10), 4065-4074.

[40] Deng, Z.-H., Luo, K.-H., & Yu, H.-L. (2014). A study of supervised term weighting scheme for sentiment analysis. Expert Systems with Applications, 41(7), 3506-3513.

[41] Dave, K., Lawrence, S., & Pennock, D. M. (2003). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. Proceedings of the 12th international conference on World Wide Web,

[42] Kang, D., & Park, Y. (2014). based measurement of customer satisfaction in mobile service: Sentiment analysis and VIKOR approach. Expert Systems with Applications, 41(4), 1041-1050.

[43] Do, H. H., Prasad, P. W., Maag, A., & Alsadoon, A. (2019). Deep learning for aspect-based sentiment analysis: a comparative review. Expert systems with applications, 118, 272-299.

[44] Nazir, A., Rao, Y., Wu, L., & Sun, L. (2020). Issues and challenges of aspect-based sentiment analysis: A comprehensive survey. IEEE Transactions on Affective Computing, 13(2), 845-863.

[45] Zhuang, L., Jing, F., & Zhu, X.-Y. (2006). Movie review mining and summarization. Proceedings of the 15th ACM international conference on information and knowledge management.

[46] Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. Marketing Science, 31(3), 521-543.

International Journal of Advances in Production Research

Entropy, Deterministic Chaos, and New Forms of Intelligibility: A Shared Frame of Reference for Physics and Psychology

William P. Fisher, Jr.[a,b,1]

[a] *BEAR Center, School of Education, University of California, Berkeley, California, USA*
[b] *Living Capital Metrics LLC, Sausalito, California, USA*
[1]wpfisherjr@livingcapitalmetrics.com

***Abstract*— Prigogine's theory of dissipative structures provides a general account of entropy-driven self-organized transitions through hierarchies of structures separated by discontinuities. The theory encompasses a wide range of evolving systems throughout nature and culture. Possibilities for operationalizing a new collective rationality spanning physics and psychology emerge from Prigogine's emphases on two distinct senses of probability, on the concept of the sufficient statistic, and on the role and limitations of the Poisson distribution in formulating a "nonlinear master equation." Unnoted by Prigogine are correspondences of all three of these issues in the mathematical foundations of statistics and measurement established in the works of Ronald Fisher and his student, Georg Rasch. The three areas of correspondence inform models enabling specifically metrological approaches to quality-assured quantification across the sciences. Prigogine's sense of "deterministic chaos" is re-expressed in measurement terms as stochastic invariance and the need for "a supplementary parameter" augmenting the Poisson distribution is related to a rating scale model of measurement. Considering these connections, this paper proposes that what Prigogine anticipates as a "new intelligibility" and a new science of "collective rationality" could be pragmatically operationalized in a new metrological infrastructure, one made coherent by the generality of entropy-driven nonequilibrium processes.**

***Keywords*— entropy; nonequilibrium processes; measurement; probabilistic models; stochastic invariance**

I. INTRODUCTION

Haynes and colleagues remark that:

> *Since the 1980s, most social science applications of entropic methods have separated themselves from intriguing but none-too-useful analogies with thermodynamics and statistical mechanics, concentrating instead on foundations in probability and statistics. We applaud this trend as a maturation of social science modeling and urge researchers to distinguish statistical entropy information from thermodynamic entropy, referring to thermodynamic ideas only when doing so is clearly necessary in ecological or nature-human interface studies.* [1] (p. 41)

One can hardly disagree that social science's entropic analogies with thermodynamics are immature. Georgescu-Roegen's 1971 book, *The Entropy Law and the Economic Process* [2], like Bailey's 1990 *Social Entropy Theory* [3], not only did not lead to a new consensus on models and methods, but difficulties encountered in substantively connecting physical entropy with evolving social and psychological systems rendered many of its uses empty metaphors [4-7]. Although the theory of entropically dissipating nonequilibrium processes is deeply rooted in the early works of Gauss, Gibbs, Fourier, and Boltzmann [8], the equilibrium conception of economic processes proved a more tractable, though more limited, integration of the nineteenth century's mutually interdependent but polarized labor theory of value and marginalism paradigms [9] - [10].

Key opportunities for demonstrating essential points of contact with theoretical and methodological fundamentals have gone unremarked, however. This has understandably obstructed abilities to imagine, plan, and obtain practical results living up to the expectations generated by the original prospects for a new science. Even so, most of the rest of the authors of chapters in the book containing the Haynes paper appear to disagree with Haynes et al.'s statement, as they all pursue the intriguing analogies Haynes disavows.

Those chapters, however, can be interpreted as making Haynes et al.'s point for them. They do so by proceeding in the usual manner from the typical assumptions of the methods employed in the social sciences, and so do not focus on substantiating important aspects of the analogy with thermodynamics. This has been the predominant manner of proceeding in a large volume of research conducted over the last 50 years extending ideas and methods involving entropy and complex systems from the natural sciences into psychology and the social sciences.

This trend began when the phenomena of noise-induced, entropy-driven physical phase transitions were found to have clear analogues in chemical and biological development and evolution [11] - [18]. This work provides a platform for imaginative explorations of a new dialogical kind of intelligibility integrating the discourses of the natural and human sciences [3], [19] - [36]. In initiating these investigations, Prigogine provides a rare example of

Nobel-level technical expertise combined with a best-selling book translated into twelve languages [37], [38]. Though much has been said in this context by way of conjectures, theorizing, and descriptive expositions, rich potentials for practical methods and results incorporating Prigogine's [39] sense of a paradigmatically "new rationality" have yet to be fulfilled.

That may change as both a general philosophical worldview and its specific mathematical technicalities are articulated in methodological terms better able to inform the conduct of a wide variety of research programs across the sciences. Given the long history of efforts failing to improve the human condition or doing so only to markedly limited extents [40], it is essential that we foreground the role of uncertainty and the potential, with the irreversible passing of time, for all conclusions to be retrospectively evaluated as premature and hasty [41], [42]. This entails an ethos of keeping the conversation that we are alive, open to new experiences, and that empowers participants in dialogue to negotiate common understandings in unique local circumstances. The goal here then is limited to some modest initial explorations of the possibility that there may be productive mathematical modeling connections to be made between nonequilibrium processes and measurement, and that these connections may be important to richer fulfillments of the potentials that many have discerned in thermodynamics for so long.

A start in this direction draws on Prigogine's theory of entropy-dissipating structures, where order increases as entropy production declines in relation to the maximum possible entropy. This current effort expands on recent research connecting entropy, information, and measurement [43] - [47]. Although that body of work is not integrated with nonequilibrium thermodynamics, it remains highly relevant to the design and implementation of metrologically traceable systems of distributed cognition [43], [44], [48] - [57].

Substantiating the analogy from thermodynamics requires a clearer understanding and operationalization of probability as an issue of central importance for characterizing social and psychological phenomena in terms of dissipative structures. Prigogine repeatedly emphasizes, but does not clarify, a distinction between two senses of probability. One of these obtains in deductive population statistics motivated by sampling problems, while the other is a sense of probability not resigned to making do with incomplete information, but which instead expresses the structure of natural law. This is a matter of central importance. After fleshing it out, Prigogine's "nonlinear master equation" is re-expressed to connect it with the form of additive conjoint, log-interval models of measurement [58] - [67] supporting quality-assured unit definitions and instrument traceability [51], [54], [68].

Prigogine's sense of the way dissipative structures of all kinds self-organize and evolve through hierarchically complex and discontinuous levels of complexity is then explicated in the context of developmental psychology. Discussion focuses on the cognitive and operational difficulties experienced in shifting individual thought and institutional policies away from automatic Cartesian assumptions of individual volition and centralized planning to the initiation of Hegelian, distributed, multilevel, embodied cognitive systems [69] - [76]. These difficulties themselves are evidence of the power exerted when individual minds are embedded in and integrated with symbolic ecologies. The fluidity of virally contagious automatic associations makes the modern Western dualistic and Cartesian worldview a formidable barrier to initiating a new unmodern nonWestern, nondualistic, Hegelian worldview embedding collective rationality in its institutional infrastructures. Thus, neither Prigogine nor Piaget fully succeeded in thinking through their basic Cartesian worldviews to transform their hidden assumptions into objects of operations at a higher order level of complexity [77], [78]. Their failures largely follow from their lack of attention to the need to solidify strategic alliances and expand networks enabling stakeholders to advance their interests more effectively via collaboration than they could in isolation [79] - [82].

Moving in this direction of ecologizing instead of modernizing [83] - [85] then entails contradicting Haynes et al.'s [1] recommendation that researchers take up "referring to thermodynamic ideas only when doing so is clearly necessary in ecological or nature-human interface studies," since the point is to cease merely describing nature and ecosystems, and to instead actively constitute human nature's integrated organism-environment units of survival, being the change we want to enact. Such lived paths substantiate developmental psychology's focus on play in environments providing ample infrastructural learning opportunities [75], [86] - [88] and correspond with the playfulness of subjective experience in language use [89] (p. 104), and with the flow of learning in the history of science [90] - [93].

Universally accessible cognitive infrastructures embedded in the social environment induce the operationalization of individuals' latent skill potentials [87], [88], [94], [95] because "organism and environment are inseparable in cognitive development" [96] (p. 646). Simply put, the question is one of augmenting individual intelligence with smart contexts [97], of obtaining "brilliant results from average people managing brilliant processes" (Cho, quoted in [98], p. 84), and recognizing that "cultural progress is the result of developmental level of support" [99] - [101]. In this context, recognizing the unity of the organism and its environment as the focus of natural selection [94], [102] - [104], it becomes apparent that common languages like the SI units constitute an infrastructural capacity essential to the advancement of science and the economy. As will be shown in the contrast of deductive probabilities and inductive likelihoods, conceiving the organism-environment as the unit of survival and adaptive evolution leads to a sense of "probabilistic epigenesis" [88] useful in imagining new possibilities for the future.

They obtain that status by enabling us to think the same thoughts in a coordinated way without having to communicate directly [105], [106] (pp. 247-257) [107] - [115]. As Hayek [116] (p. 88) [117] put it,

> *The problem is precisely how to extend the span of our utilization of resources beyond the span of the control of any one mind; and, therefore, how to dispense with the need of conscious control and how to provide inducements which will make individuals do the desirable things without anyone having to tell them what to do.*

Hayek's specification of inducements will prove to be key in addressing this problem of how to embed a collective rationality in the infrastructural cognitive supports provided by the external social environment. Prigogine [118] (p. 504) similarly remarks on how models of nonequilibrium processes surpass the capacities of traditional equilibrium models in economics, saying that:

> *The interest of this class of models is that they enable us to make the interplay between the actors and the constraints of the environment more transparent…. Let us emphasize the importance of such models for social sciences in order to make the decision mechanisms more transparent in a democratic society: we have here an example of a process of evolution in which science and collective rationality may interact in a constructive way. Perhaps this will be a way of demythologizing the process of collective decision making, without negating its complexity. Models alone will of course not be a substitute for political decision making, but they may help to make their implications more explicit.*

The models and "new laws of nature" Prigogine [119] holds as emerging at "the end of certainty" require recognizing, accepting, and integrating the uncertainty entailed by evolving entropy-dissipating systems. In these systems, three discontinuous levels of complexity are simultaneously enacted: a concrete within-individual micro level, an abstract individual meso level, and a formal population macro level. Prigogine and Stengers [38] (p. 300) say:

> *…an essential characteristic of our scheme is that it does not suppose any fundamental mode of description; each level of description is implied by another and implies the other. We need a multiplicity of levels that are all connected, none of which may have a claim to preeminence.*

As a result of his ethnographic study of experimentalists, metrologists, and theoreticians in microphysics, Galison [120] (p. 143) similarly proposes an "open-ended model" that is "tripartite in allowing partial autonomy to instrumentation, experimentation, and theory," and that leads us to "expect a rough parity among the strata—no one level is privileged, no one subculture has the special position of narrating the right development of the field or serving as the reduction basis." Galison [121] (p. 46) explains, remarking on how "representing meaning as locally convergent and globally divergent seems paradoxical" while at the same time recognizing that:

> *It seems to be a part of our general linguistic ability to set broader meanings aside while regularizing different lexical, syntactic, and phonological elements to serve a local communicative function. So too does it seem in the assembly of meanings, practices, and theories within physics.* [121] (p. 49)

That is, everyday language does not typically use abstract phonemic and grammatical standards to impose uniform idealized conceptions on unique local circumstances. Instead, arbitrary abstractions set up as consensus standards serve as the media through which shared understandings are negotiated by relating multifaceted ideas to concrete things. In everyday language, the semantic triangle of ideas, words, and things functions as an assemblage packaging a heterogeneous array of structures, functions, and fluctuating circumstances in intuitively accessible linguistic systems.

Semiotics has then expanded into ecosemiotics, biosemiotics, and cybersemiotics as the role of language as the irreducibly complex vehicle of thought has focused interest on transdisciplinary models of science [122] - [129]. Analogous senses of multilevel complex assemblages have been explored and documented by other investigations in the history of science emphasizing the roles of measurement standards and instruments as mediating social relations, theory, and data [81], [82], [106], [130] - [152]. Though multilevel complexity is difficult to grasp conceptually, everyday language use and repeated daily measurements of time, distance, temperature, etc. involve practical experience in how it works.

In like fashion, in psychometrics, Guttman [153] (pp. 79-81) also emphasizes that three levels are employed simultaneously when he distinguishes informal everyday language from formal technical terminology and from mathematical symbolization, en route to articulating a theory and practice of cumulative science. Each of Guttman's levels can be seen as implying semiotic distinctions between levels for concrete things and data, abstract words and instruments, and formal ideas and theories within them. Wright [154], another psychometrician, explicitly expands on the levels of complexity offered in Peirce's semiotics to describe the evolution of understanding in science and measurement, while Stenner and Horabin [155] describe the evolution of construct theories as progressing from intuitions to data to prediction. And to underscore the multilevel functioning of cognitive operations, developmental psychology recognizes that people utilize multiple levels of cognitive complexity at the same time, and typically employ the lowest level of performance allowed by the environment [96], [156].

The uncertainties entailed by the chaotic dynamics characterizing these levels of complexity set limits on the extent to which the interplay between the individual and its environment can be made more transparent. That transparency will never be perfect, but much can be done to improve on the obscurities and confusions dominating today's systems by recognizing, accepting, and acting on the all-pervasive ubiquity of uncertainty:

> *...even in classical physics we get randomness and unpredictability.... pure mathematics, in fact even elementary number theory, the arithmetic of the natural numbers, 1, 2, 3, 4, 5, is in the same boat. ... So if a new paradigm is emerging, randomness is at the heart of it.* [157] (p. 13)

We are well past the point in time at which the historic importance of uncertainty as a paradigm-setting principle [157] - [163] ought to have infused every level of discourse in every field of human endeavor. Given that his proofs on the irreducible incompleteness and inconsistency of self-referential systems, and his "discovery that there are arithmetical truths which cannot be demonstrated formally" [164] (p. 101), have had little effect on the practices of mathematicians, it is not surprising that Gödel wrote a letter to his mother expressing his disappointment that his work had not impacted mathematics the way that his friend Einstein's work had impacted physics [157] (p. 13). Though it will have taken much longer to arrive, Gödel's day may yet come. Gödel's proofs inspire confidence in the validity and truth of the shift in the conception and practice of scientific method entailed by the foregrounding of uncertainty advocated here [165] (p. 52).

That confidence will be needed as the limits on knowledge become even more complex. It will be imperative that we move beyond systematic integrations of concrete, abstract, and formal levels to yet higher orders of complexity. We tend to devalue and brush aside paradoxical complications, ambiguities and inconsistencies in daily life, such as the contradictions obtaining in obeying or disobeying a parental command to stop acting like a child, or those involved in taking the initiative while not crossing boundaries.

In this vein, contrary to Kuhn's [166] emphasis on convergent thinking in coherent paradigms, Prigogine and Stengers [38] (pp. 307-309) take issue with Kuhn's sense of normal science, emphasizing the way undercurrents of nagging discrepancies persistently appear, are minimized or ignored and disappear, and re-appear in unexpected contexts, in the history of science. Similarly, in a complementary fashion, others have shown healthy scientific fields to be productive as a result of perspicacious divergence in thinking, where disagreements provoke learning, investments in demonstrating proofs, and developments toward unexpected insights [121] (pp. 46-49, 843-844) [167] - [170]. Recognizing that universally uniform consensus on fully determined facts has never been and never will be achieved, Prigogine and Stengers [38] (p. 299) conclusively assert that "the epoch of certainties and absolute oppositions is over." Indeed, nothing is more certain than uncertainty [171].

Agreeing to disagree in productive ways becomes ever more important in the contexts defined by systems of systems at the metasystematic level of complexity, by supersystems at the paradigmatic level, and by systems of supersystems at the cross-paradigmatic level [90], [91], [172] - [174]. The provisional clarifications that might be achieved by explicitly addressing uncertainty and modeling it in practically applicable stochastic formulations might lead toward empowering transformations [85] of today's disempowering organizational paradoxes [175].

We here intentionally seek to define and create institution-level infrastructures for ecologized economies of thought at a paradigmatic level [176]. This intention is a continuation of Prigogine's [15] (p. 12) introduction of mathematical methods for characterizing "a whole hierarchy of structures separated by discontinuities" enabling "a concrete, unified description of the macro-world," where "the concept of stability really reconciles the unity of laws with the existence of well-defined levels of description." These methods must then integrate individual mental operations with the external environment's infrastructural supports at multiple discontinuous levels of complexity, in a manner replicating today's existing integrations of scientific, legal, market, and communications networks and standards [108] - [111], [177] - [181]. In so doing, we aspire to fulfilling Prigogine's goal of a new collective rationality.

II. DISTINGUISHING PROBABILITY FROM LIKELIHOOD

Prigogine [17] distinguishes between two kinds of probability while arguing that transitions through "a hierarchy of structures separated by discontinuities" are entropically driven by spontaneous reorganizations of functions—fluctuations—and that "this conception is applicable to a large number of situations, including the functioning of cognitive structures envisioned by Piaget" [17] (p. 263) [19], [39], [118], [182] - [187]. Brooks and Wiley [11] (p. 356) concur, saying, "The second law [of thermodynamics] is thus more than the natural law of energy flows; it is the natural law of history" [25]. Prigogine [17] (p. 263) [16], [38], [119] concludes by saying, "We are perhaps moving towards a new discipline which will inherit from physics the cares of the world, of quantitative description, and from classical metaphysics the ambition of finding a coherent global image that would include man."

At the physics end of this new discipline, Prigogine argues that, from his perspective, in quantum theory "the basic quantity is no longer the wave function corresponding to a probability amplitude, but probability itself. …Probability is no longer a state of mind due to our ignorance, but the result of the laws of nature" [119] (p. 132), [17] (p. 262), [188] (p. 5). These laws comprise a "new form of intelligibility as expressed by irreducible probabilistic representations … [that] deal with the possibility of events, but do not reduce these events to deductible, predictable consequences" [119] (p. 189).

Prigogine’s sense of a different, nondeductive kind of probability directly expressing laws of nature independently reproduces the inductive likelihood functions Ronald Fisher [189], [190] contrasted with deductive population statistics’ probabilities. Fisher [189] (p. 367) wrote:

> *The conclusion is drawn that two radically distinct concepts, both of importance in influencing our judgment, have been confused under the single name of probability. It is proposed to use the term likelihood to designate the state of our information with respect to the parameters of hypothetical populations, and it is shown that the quantitative measure of likelihood does not obey the mathematical laws of probability.*

Duncan and Stenbeck [191] (pp. 24-25) argue that this contrast of likelihood and probability marks an essential difference between scientific and statistical models:

> *The main point to emphasize here is that the postulate of probabilistic response must be clearly distinguished in both concept and research design from the stochastic variation of data that arises from random sampling of a heterogeneous population. The distinction is completely blurred in our conventional statistical training and practice of data analysis, wherein the stochastic aspects of the statistical model are most easily justified by the idea of sampling from a population distribution. We seldom stop to wonder if sampling is the only reason for making the model stochastic. The perverse consequence of doing good statistics is, therefore, to suppress curiosity about the actual processes that generate the data.*

Duncan and Stenbeck [191] (p. 23) [192] say they emphasize this distinction between statistical models and scientific models “with all the rhetorical force we can muster.” A large chorus of others writing before and since [67], [193] - [203] expand on this theme. Guttman [204] (p. 329), for instance, states that “Measurement theory…deals with the construction of structural hypotheses rather than with inference from samples.” Statistical models describe population-level data distributions to support deductive inferences from samples. In this paradigm, models are fit to data, prioritizing the maximization of explained variance or the minimization of p-values in significance tests. Model-data mismatches are rectified by modifying the model, often via the addition of multivariate interactions. The hypothesis of a reproducible unit quantity that remains invariant across instruments and samples is not formulated or tested.

Scientific models, in contrast, prescribe the form of individual level response functions to inductively infer the definition and comparison of generally applicable unit quantities bound by uncertainty. In this paradigm, data are fit to models specifying the univariate structure of quantitative measurements. Model-data mismatches are rectified by modifying the data, not just in the sense of the particular observations included in tests of the quantitative hypothesis but more fundamentally in the sense of attending to the content of the questions asked.

Ronald Fisher [189] developed the mathematical basis for the contrast between (a) scientific measurements, where individual observations comprise a sufficient statistic (a count or sum score) extracting all available information from the data, and (b) statistical analyses in which group-level means and standard deviations are sufficient for reproducing normal distributions. These concepts and distinctions remain today just as confused and blurred as they ever have been. Explicit mention of the differences between deductive sampling probabilities and inductive response likelihoods is lacking in most examinations of probability in measurement modeling. When they are mentioned, the differences are minimized.

For instance, a recent text on metrological infrastructure [205] (pp. 46-47) notes that two mutually inconsistent versions of probability co-exist in the GUM (Guide to the Expression of Uncertainty in Measurement) [206]. One version involves a “conventional” focus on event frequencies, and the other, subjective degrees of belief. Though not referred to as such, this distinction plainly involves the difference between deductive inferences motivated by sampling problems and inductive inferences motivated by response processes. This internal inconsistency is said to be “harmless” and that, when a decision between them must be made, the GUM endorses the conventional view.

In another example [207] (pp. 39-40) in which the differences between inductive likelihood functions and deductive sampling probabilities are noted, it is recognized that “the likelihood function is not a probability distribution” and that “maximum-likelihood estimation is not a probabilistic estimation.” But the overriding assumption of probability as singularly involving sampling problems leads to the dismissal of the likelihood function as ”not fully convincing” and largely irrelevant to measurement modeling. When inductive inference problems are addressed, they are typically taken up in Bayesian terms [207] (pp. 48-54), and in the course of identifying the value of the measurand from the measurement system’s observed output (i.e., during a restitution or reconstruction process) [208], [209].

The inferential connection made in restituting the measurement value in relation to an SI unit standard suggests a potentially productive point of entry for expanding the conceptualization and operationalization of inductive inferences and response process likelihoods into measurement systems featuring quality-assured metrological traceability [51] (p. 48), [43], [68] (pp. 71-74). An influential source of possibilities for that kind of an expanded appreciation of likelihood functions in metrology are found in the works of Rasch [63], [65], [210]. After studying in 1934-1935 with Ronald Fisher in London, and with the econometricians Frisch and Tinbergen in Oslo, Rasch formulated probabilistic models for measurement blending Fisher’s emphasis on the sufficiency of individual-level response processes with Frisch and Tinbergen’s application of Maxwell’s method of analogy [111]. Rasch

[63] (pp. 110-115) structured the form of his models via an analogy from Maxwell's analysis of Newton's Second Law, incorporating the form of lawful regularity into the model parameterization.

In taking this approach, Rasch's measurement modeling concepts stand in complete opposition to the models and methods associated with Item Response Theory (IRT) [211], as has been repeatedly asserted by all of the major proponents contributing to the advancement of Rasch's perspective [67], [193], [201] - [203], [212] - [218]. As Linacre [219] (p. 926) put it, "The Rasch dichotomous model is a derivation from measurement axioms. It has nothing to do with the normal ogive model" used in IRT. Even IRT advocates admit that "The [IRT] theta-scale, or any linear transformation of it, however, does not possess the properties of a ratio or interval scale, although it is popular and reasonable to assume that the theta-scale has equal-interval properties" [220] (p. 87).

Going along with the misconceived association of Rasch's perspective on measurement modeling with IRT simply because it is "popular and reasonable" amounts to nothing more than fallaciously appealing to expert authority, begging the question, equivocating, and appealing to popular opinion, all of which unscientifically assume something is true in the absence of a logical argument or evidence [221] (p. 20).

Rasch's principled perspective on designing instruments with the intention of empirically fulfilling explanatory ideals to fit-for-purpose quantitative uncertainty tolerances accords with

- Kuhn's [222] (p. 219) historically informed perspective that "The road from scientific law to scientific measurement can rarely be traveled in the reverse direction;"
- Butterfield's [223] (pp. 16-17, 26, 96) sense of scientific advances not being based in primarily empirical accumulations of observed data but requiring a geometrical "thinking cap;" and with
- Kant's [224] (p. 20) perspective that science does not follow "nature's leading strings," but compels nature to answer questions of reasoning's own determination in an unfolding dialogue.

Appropriating without citation Tinbergen's application of Maxwell's method of analogy [111], [225], [226], Rasch conceived of measurement as a probabilistic projection of the form of a scientific law, using it as a means of seeing whether experience might be amenable to that kind of organization. As Rasch wrote:

> *...the acceleration of a body cannot be determined; the observation of it is admittedly liable to ... 'errors of measurement', but ... this admittance is paramount to defining the acceleration per se as a parameter in a probability distribution -- e.g., the mean value of a Gaussian distribution -- and it is such parameters, not the observed estimates, which are assumed to follow the multiplicative law [acceleration = force / mass]."*
>
> *Where this law can be applied it provides a principle of measurement on a ratio scale of both stimulus parameters and object parameters, the conceptual status of which is comparable to that of measuring mass and force. Thus, ... the reading accuracy of a child ... can be measured with the same kind of objectivity as we may tell its weight* [63] (p. 115).

Rasch here built on Ronald Fisher's [103] (pp. 103-104) point that, "because all laws of natural causation were essentially laws of probability, the predictability of a system has the same basis in the natural as in the social sciences" [227] (p. 289). In accord with Prigogine's sense of "deterministic chaos," Fisher held that, "far from being contradictory, the notions of probability and determinism are intrinsically related" [227] (p. 290). The key distinction between likelihood and probability is maintained here in that the Gaussian distribution referred to by Rasch concerns the errors of measurement associated with each individual estimate, whether of stimulus or response parameters. All further references here to probabilistic models of measurement assume this distinction is understood.

Wright [228] (p. 32), the foremost proponent of Rasch's ideas from the 1960s through the 1990s [229], [230], accordingly says that what Rasch accomplished is "a definition of measurement, a law of measurement. Indeed, it's *the* law of measurement." Probabilistic conjoint models [59], [61], [231], [232] operationalizing this law of measurement have informed several decades of research and practice in the development and deployment of quantitative systems for high stakes admissions, graduation, certification, and licensure contexts [233], [234], in classroom applications facilitating formative feedback [235] - [241], in the context of Piagetian developmental psychology [242] - [250], in adaptive and AI instrument adminstrations [251] - [253], and others involving tens of millions of measurements annually [254] - [258], in education, health care, sustainability, and other fields.

Research in the domain of dissipative structures, however, has not incorporated the structure of natural law into the form and function of measuring instruments calibrated to defined unit quantities. Prigogine stresses the empirical, phenomenological groundedness of the new sense of probability he has identified, and notes that it is not the same thing as the probabilities deduced in accommodating the limits on knowledge imposed by statistical sampling problems. But Prigogine's research and that of others applying his ideas does not specifically model and evaluate invariant structural dimensions of collective action projected by and estimated from individual observations. Neither are these experiments conducted with the aim of calibrating tools intended for distribution throughout networks of end users whose decisions and behaviors constitute the collective rationality that was modeled.

On the contrary, the probabilistic methods implemented in applications of the theory of dissipative structures are formulated and applied in a deductive and descriptive statistical manner that neither identifies nor calibrates the hypothesized collective dimension on the basis of individual response likelihoods. The promised laws of nature are confirmed only in summary statistics describing processes from the outside in and top down, not in distributed measurements inscribing processes from the inside out and bottom up. Not being formulated in terms informing active participation in modeled processes on mass scales, the statistically described collective rationality has no medium of implementation. We now turn to how that kind of formulation might be achieved.

III. RE-EXPRESSING PRIGOGINE'S "NONLINEAR MASTER EQUATION" AS A PROBABILISTIC MODEL OF MEASUREMENT

Prigogine's [16] (pp. 115-119) [188] (pp. 28-34) "nonlinear master equation" can be restated as a probabilistic model of measurement, one that "belong[s] to the same class that metrologists consider paradigmatic of measurement" [50] and that provides "a specifically metrological approach to human-based measurement" [54] (p. 26). The potential for a meaningful restatement of the master equation follows from Prigogine's distinction between empirically descriptive phenomenological equations and theoretically prescriptive master equations, which then provide a model of a natural law. Prigogine's emphasis on individual-level processes is substantiated in terms of the same Poisson and Markov processes employed in probabilistic measurement. Prigogine stakes the objectivity of the measurements made on the demonstrable generality in nature of Poincare resonances, which are taken as the basis of a structural analogy to cognitive and social phenomena.

A. Poisson Limitations and the Need for a "Supplementary Parameter"

Prigogine [16] (pp. 116-117) remarks on shortcomings of the Poisson distribution limiting its relevance to stable systems. This limitation makes it necessary to bring in a "supplementary parameter scaling the extension of the fluctuations" to make it apply in unstable systems [16] (p. 117). Fluctuations compromising the stability of the macroscopic state may provoke a transition to a different level in a discontinuous hierarchy. But the appearance of a physical instability depends on fluctuations of a critical size, as they behave differently at different size spatial extensions: "only fluctuations of *sufficiently* small dimensions obey Poisson statistics" [16] (p. 117; emphasis added). Prigogine then remarks that "this is a very important result because it implies that, conversely, only fluctuations of a *sufficient* extension can attain enough importance to compromise the stability of the macroscopic state considered" [16] (p. 117; emphasis added).

Prigogine and Allen [188] contrast the applicability of the Poisson distribution and the law of large numbers in phenomenological equations for systems near equilibrium with the lack of an identifiable relation between them in nonlinear systems far from equilibrium. They say:

> *...the identification between these macroscopic equations and the first moment equations may become completely incorrect, as the probability distribution may cease to be sharply peaked, in which case the law of large numbers no longer applies. The probability distribution may, for example, become double humped, in which case the average value of a variable and its variance become meaningless quantities, since they completely fail to define the double-humped curve* [188] (pp. 26-27).

Though no mention is made of the concept of statistical sufficiency, Fisher's [189] proof of the meaninglessness of the average and its variance outside of the context of a normal Gaussian curve is clearly invoked. Also implicated is the idea of identified models, developed in econometrics by Frisch, Koopmans, and Reiersøl as a criterion for the practical applicability of models intended to guide real-world implementations [259], [260]. Though Rasch makes no mention of model identifiability, Koopmans and Reiersol thank him for his input in "fruitful" discussions, and he effectively applied the concept via his emphasis on sufficiency [111], [216].

The nonlinear transitions of concern correspond in psychology to changes from (a) counts of one kind of process, or a category of ratings accumulated across observations of performances within a particular level of difficulty, to (b) counts of another kind of process, or ratings concerning performances at a higher level of difficulty, where what Prigogine [16] (p. 117) refers to as the "supplementary parameter" represents the thresholds of the transitions across levels. Models incorporating this kind of a parameter are widely used in assigning partial credit in the scoring of learning outcomes, development and growth, and in rating performances, attitudes, behaviors, and skills [261] - [264]. As suggested in making the distinction between scientific and statistical models, a matter of fundamental importance in both Prigogine's theory of dissipative structures and measurement theory concerns the sufficiency of the transitions from one category to another.

As he stated in an autobiographical passage, Rasch [65] (p. 66) did not appreciate the scope of what he had accomplished in formulating the measurement model associated with his name until he was provoked by Frisch to see that he had answered the question:

> *Which class of probability models has the property in common with the Multiplicative Poisson Model, that one set of parameters can be eliminated by means of conditional probabilities while attention is concentrated on the other set, and vice versa?*

The inferential separation of parameters in this class of probability models allows person or object measurements to be compared independent of the particular questions or indicators used as the medium of observation [264] - [266] [267] (pp. 52-55). Some years after Frisch's astonishment at the "disappearing parameter," Rasch [210] (pp.

104-105) [65] formulated the epistemological concept of specific objectivity to characterize the quality of the inferences and comparisons supported by measurements modeled in this way.

Rasch intentionally structured his models in the form of Fisher-Darmois-Koopman-Pitman additive exponential models because these admit the observed score summed from Poisson counts of correct answers or performance/frequency/attitude ratings as a sufficient statistic for the individual-level response process likelihoods [65], [268] - [277]. Prigogine made use of these distributions for much the same purpose, but without structuring the model in the form of Maxwell’s analysis of Newton’s second law, as Rasch did.

What Prigogine calls “noise-induced phase transitions” are shifts from one categorical state of existence, cognition, moral bearing, mode of being, form of life, or species, to another. These take place as the uncertainty associated with being at a given level increases, the likelihood of staying at that level decreases, and the likelihood of a shift to a higher level increases. This is what happens in sociocognitive measurement when development away from one level makes a person’s status increasingly ambiguous before it settles into a new stability at a higher level.

Prigogine and Allen [188] (pp. 24-28) give a simple example of

> *…how the probabilistic and phenomenological approaches are linked and how in the description of a nonlinear system whose basic mechanisms define a Markov process, the detailed study of the behavior of fluctuations can give an important new insight into the phenomena encountered.*

The simple example given notes that the Poisson distribution and the law of large numbers apply as long as the system is stable and the probability (likelihood) is peaked around a single value, but as a bifurcation point is approached, where the dominance of one form of life (species, culture, technology, ability level, health status, etc.) declines in relation to the increase of another, the variance diverges and the distribution spreads into a double-peaked probability curve. But the variance diverges in a way that leads to “a new type of law of large numbers,” where the overall probability distribution is flattened, and “we can trace the line of coexistence between the two solutions on the basis of the stochastic potential” [188] (p. 27). This new type of law of large numbers corresponds to a rating scale model of measurement [261] - [264].

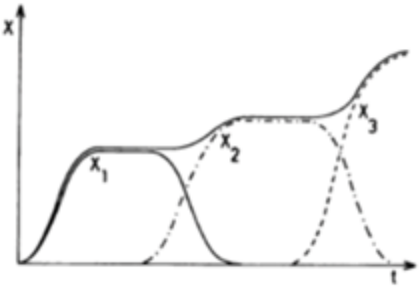


Figure 1. Evolutionary succession of species in the same niche [188] (Figure 1.15, p. 30).

This correspondence is partially illustrated by Prigogine and Allen’s [188] (p. 30) Figure 1.15 (Figure 1 here) and rating scale category transition likelihood curves (Figure 2). (Though typically referred to as probability curves, in an effort at maintaining a more rigorous distinction between probabilities and likelihoods, they will be called likelihood curves here.) Each set of curves in Figure 2 informs an interpretation of a single person’s (or any group’s mean) measurement in relation to a single test, assessment, or survey item. Person measurements lower than an item’s calibration are to the left of center in each figure, and measurements higher than the calibrations are to the right. The further to the left a measurement is, the less likely a fully correct or highly agreeable response is, as the question asked is more difficult than the person is able. Conversely, as measurements increase, moving horizontally further to the right, the more likely successful or agreeable responses become. As shown in the dichotomous situation at the bottom right in Figure 2, involving responses scored in two categories, when the person ability is the same as the item difficulty, the odds of a correct or agreeable response are 50-50.

To take another example, in the curves at the upper left in Figure 2, someone with a measurement of about ¬0.5 logits has about a 0.75 likelihood of success at level 1, with a 0.125 likelihood of both failure (0) and greater success (2), and a negligible chance of the highest level of success (3). This person’s very low chance of success on the more difficult variation on that task may be of no consequence if the task is irrelevant to the demands of this person’s daily life. With further development in this person’s abilities, however, the task’s relevance increases, and so does the person’s ability to successfully address it. At some point around 1.2 logits on the (horizontal) ability scale in the upper left curves in Figure 2, the person’s ability to succeed on the easier level 1 version of the task will be just as likely as their ability to succeed on the harder level 2 version. Multiple test, assessment, or survey items, as well as the challenges of everyday life, targeting this level are likely to be answered correctly or rated successful about half the time, which is to say, randomly. The person’s location in relation to the two versions of the task is unstable, flipping back and forth between them.

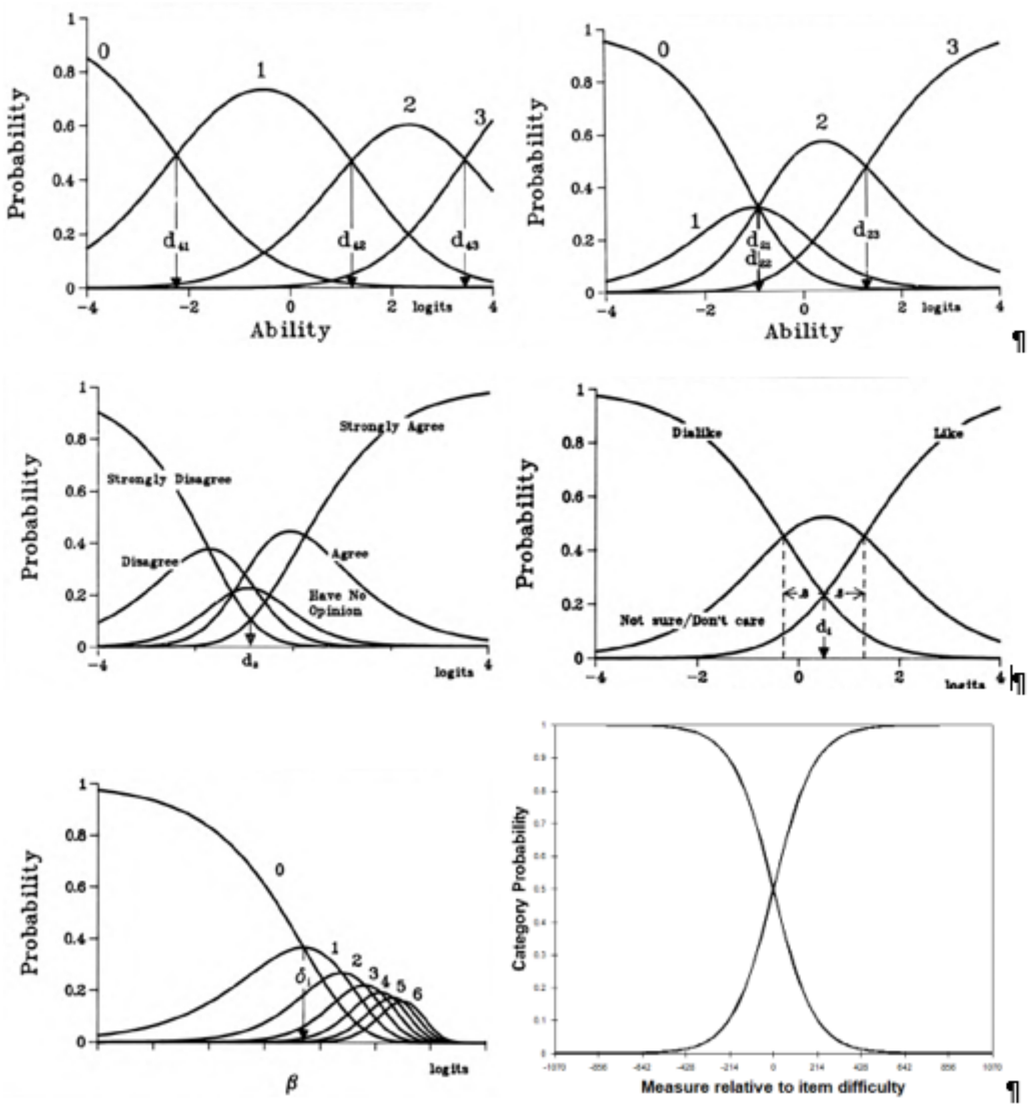


Figure 2. Examples of response category probability (likelihood) curves in measurement theory [264], [278].

This is the level at which the person's status reaches a bifurcation and is fluctuating. If the existing environment offers clear supports inducing behaviors at the higher order level, a transition to that level becomes more likely than if individuals must create those supports themselves or do without them. This is the situation of "probabilistic epigenesis," where individual response likelihoods are damped or amplified by interactions between internal capacities and environmental features [88] (pp. 52-53), [279] - [281]. Though analogies between the genotype and the theory of the measured construct, and between the phenotype and the physically material instrument, may be apt, evolving mutability is not a strict function of random fluctuations but depends on the expression of capabilities finding traction in the external environment. Far from being "a simple additive function of gene x environment interactions," advances in epigenetics and genomics had not coalesced around a core set of principles characteristic of a paradigm in either biology or psychology by 2015 [88] (p. 10), and arguably still have not. Methodical distinctions between probability and likelihood situated in relation to Prigogine's theory of dissipative structures, models for measurement, and developmental psychology might contribute to the coherence of such a paradigm.

The transition thresholds in Figure 2 could be made analogous to Figure 1's [188] (p. 30) illustration of the succession of species in the same niche by making the vertical axis in Figure 1 represent likelihoods summing to 1.0 across the curves. The shifts across these ranges of hills occur as the likelihood of one cognitive or biological form of life dominating decreases and that of another increases.

Figure 1 shows each successive curve as higher than the preceding one, but this need not always occur, especially if the characteristics of the ecological niche change. The analogy here is complemented by the way these likelihood functions are akin to resonating clouds of matter, energy, and/or information. The model posits the infinite populations of all possible persons and items participating in the measured construct. The large numbers of actual people and tasks participating in the construct in daily life are sampled in experiments modeling their interactions. This modeling process has the aim of creating tools embodying people's collective rationality in ways useful for defining general paths of least resistance adaptable to individual circumstances. The experimental process replicates real life and simplifies it to an actionable model that can be exported from the laboratory back into the world. Care must be taken to separate and balance the unrealistic ideal of the heuristic fiction posited by the formal law structuring the model in relation to the arbitrary medium of the functional measurement's abstract standard unit and in relation to the chaotic randomness and fluctuations of individuals' concrete data. The model becomes a shared medium for communication to the extent that it is augmented by quality assured traceability, consensus standards, laboratory accreditation, etc. contextualizing the management of individual learning, healing, and growth.

B. Modeling stochastic resonances and the microprocessing of facts

The resonating clouds of stochastically appearing and disappearing data points provide a macroscopic analogue of the Poincaré resonances Prigogine [119] (p. 151) says "measure themselves" by means of their interactions. The explicit, overt measurements made in laboratory research extend the implicit, latent measurements being made in nature, and in society. The models are useful in applications only insofar as they in fact have successfully leveraged their sufficient statistics and identifiability to represent real phenomena. The structures, processes, and outcomes of the real world are encapsulated in laboratory models useful to the extent they can inform decisions and behaviors in everyday life.

When the limiting conditions in which repeatable references to objective facts are understood and have been reproduced often enough in varying circumstances to inspire confident applications, there is no longer any need to continue the "microprocessing" of those facts, and the model can be packaged in an intuitive technology for commercialized export from the laboratory [130] (pp. 143-144), [137] (pp. 132-135). In examples from educational psychology, this process has repeatedly been effectively implemented [52], [53], [254], [257], [258], [282] - [284].

Prigogine argues that Poincaré resonances are modeled via the theory of dissipative structures as objectively repeatable and reproducible phenomena in a way that removes the need for the "mysterious intervention" of an observer at the quantum level [119] (p. 131), [285]. As Prigogine [119] (p. 46) puts it, "Instability driven by resonances plays a fundamental role in changing the formulation of quantum mechanics." Alternative approaches to eliminating the "extravagant role in the evolution of nature" [119] (p. 15) assigned the observer by Bohr's Copenhagen interpretation of the quantum phenomenon are also offered by Bohm's [286] - [300] ontological interpretation and by Wheeler's [301] - [306] sense of a participatory universe. Fuller elaborations of Bohm's explicitly and Wheeler's implicitly Hegelian perspective on these issues are in development.

Instability driven by resonances also played a fundamental role in changing the formulation of social and psychological measurement. These resonances explain why "the stochastic model of Rasch, which might be said to involve weaker assumptions than Guttman uses [in his deterministic models], actually leads to a stronger measurement model" [307] (p. 220). An analogy between the physical phenomenon of stochastic resonance and response process likelihoods in psychological measurement [308] - [310] [311] (pp. 69-71) may be substantiated by dissipative structures' incorporation of Poincaré resonances, which Prigogine [119] (pp. 36, 131) describes as playing a key role in deterministically chaotic nonequilibrium systems [285], [312] - [314].

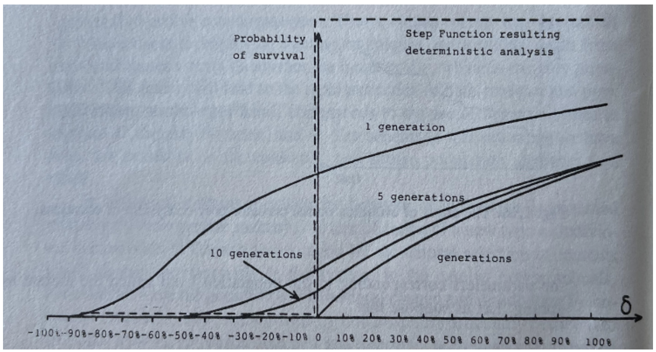


Figure 3. Deterministic and probabilistic survival likelihoods [188] (Figure 1.17, p. 34).

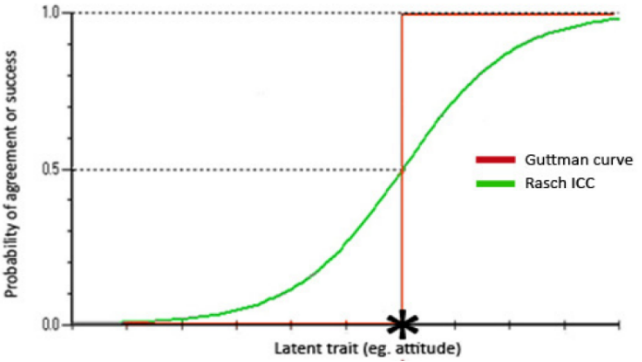


Figure 4. Deterministic Guttman and probabilistic Rasch response functions [337] (p. 253).

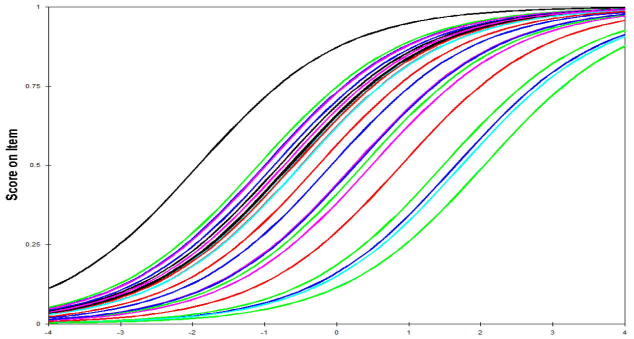


Figure 5. All of a single instrument's item characteristic curves [278].

Prigogine's focus on noise-induced phase transitions is also relevant in the context of measurement modeling for the value obtained for estimation and substantive theorizing from the focus on individual response likelihoods. The gradient of shifting likelihoods is associated with an information function that operates in the etymological root sense of "inform." That is, a strictly deterministic step function of the kind described in Prigogine and Allen's Figure 1.17 and by Guttman's [315] scalogram approach (see Figures 3 and 4) results in an "attenuation paradox" [316] - [318]. The paradox emerges as transitions between levels become more deterministically defined, and more certain, which then diminishes the capacity to estimate the distances between them.

Note that attention to individual response likelihoods in the measurement modeling context contrasts with the statistical modeling approach to the attenuation paradox taken by Lord [319] (pp. 4-5), where "a test score is a sample statistic that is to be used to test some hypothesis." The statistical focus on group-level statistics emphasizes uses of p-values and explained variance in evaluations of the fit of models to data [320], which leads to methods and results paradigmatically opposed to those developed using the measurement focus on individual-level response processes emphasizing meaningful interpretations in evaluations of the fit of data to models [67], [193], [201], [202], [321].

The deterministic structure's vertical step function contrasts with the additional information provided by the smoothly transitioning likelihoods illustrated by the generational survival and item characteristic curves in Figures 3, 4, and 5. Regarding Figure 3, Prigogine and Allen [188] (p. 33) state a model of a probability of surviving long enough on average to reproduce as:

$$P_{\text{surviving to reproduce}} = \delta \,/\, 1 + \delta - e^{-\delta \,/\, 1 + \delta} \qquad (1)$$

Here, δ is the fractional change in selective advantage offered by a biological mutation, or, extrapolating via Prigogine's analogies, a newly achieved Piagetian level of development, or a social innovation. Expanding δ into two terms capturing both the mutation and the context of its fractional change in selective advantage, and rewriting the equation as a probabilistic conjoint measurement model [61], [282], this becomes:

$$P_{\text{surviving to reproduce}} = e^{(\beta - \delta)} \,/\, 1 + e^{(\beta - \delta)} \qquad (2)$$

which gauges reproductive success as a function of the difference between the new form of life's survivability β and the challenges δ posed by the environment. Levels of additional advantages represented by what Prigogine refers to as a "supplementary parameter" can be modeled in the form of:

$$P_{\text{surviving to reproduce}} = e^{(\beta - \delta - \tau)} \,/\, 1 + e^{(\beta - \delta - \tau)} \qquad (3)$$

where τ [262] - [264] represents the added difficulty posed by challenges within specific ecosystem niches, or, conversely, the added ability exhibited by a mutation or innovation. In a further expansion:

$$P_{\text{surviving to reproduce}} = e^{(\beta - \delta - \tau - \lambda)} / 1 + e^{(\beta - \delta - \tau - \lambda)} \qquad (4)$$

where λ could be another source of added environmental challenges that vary depending on the context, such as a predator or a systematic lack of resources in an ecological niche, a judge rating performances, a multicomponent task, an explanatory scheme predicting item locations, etc. [322] - [329]. Additional model features can be added to address problems of items' or raters' local dependencies, systematic differential item functioning, latent class variations, etc. [330] - [336].

The Socratic midwifery of a fully formed mode of being able to take on a life of its own in an appropriately resourced environment depends, in Prigogine's language, on the size of the fluctuation. In the substantive terms of psychological measurement, it depends on the likelihood of achieving a decisive consolidation in an ability or performance at a sustainable new level. Sustainable cognitive operations and behaviors are functions of uncertainty. At the bifurcation point—the category transition threshold—achievement may fluctuate across tasks posing equivalent challenges with 50-50 odds, randomly. With low numbers (fewer than ten) of opportunities to demonstrate the level of achievement, performances that average to random coin tosses may vary in their individual likelihoods much more widely than those based on larger numbers of observations (30 or more). Either way, successes will default to the more likely, previous level of achievement until a new stability coheres with higher likelihoods and less uncertainty at the new level.

Prigogine and Allen [188] (p. 27) say that in large physicochemical systems there is always one maximum likelihood that is dominant, but this is not the case in sociocognitive systems, as is illustrated at the upper right and middle left in Figure 2. Even in large datasets categorizations may be too finely graded to permit every distinction to have the highest probability (likelihood) within even a narrow range. In addition, an instrument administered to a sample that performs at markedly high or low levels, and so is off target, may have too few responses at the unoccupied extreme to allow estimation of the transition thresholds. This situation raises the problem of calibrating instruments on appropriately chosen samples, and not on samples drawn for experimental or applied purposes. Items included on calibrated instruments should be anchored at standard unit values, not recalibrated to a different unit in every new application. Disagreements as to whether it is reasonable to collapse unobserved, and so disordered, categorical distinctions on the basis of the data, or if they should be retained on the basis of the construct theory, has led to some controversy around the value of categories and their scoring [338] - [342]. Additional issues concern the summing of graded response models' cumulative likelihoods, which are always ordered but which then prevent the inferential separation of item and person parameters [343], [344] (pp. 322-324), [345], [346].

Prigogine and Allen [188] (pp. 26-27) distinguish between stable linear systems exhibiting empirical (phenomenological) proportionality and unstable nonlinear systems far from equilibrium "in which the kinetic equations give rise to bifurcations" and "the identification between these macroscopic equations and the first moment equations may become completely incorrect, as the probability distribution may cease to be sharply peaked." When this kind of a double-humped distribution occurs, the item's error distribution becomes bimodal, rendering the location of the transition threshold on the scale unresolvable [347]. Whether this problem should be resolved by anchoring the thresholds at established values, equating with a previous sample, revising a poorly designed rating scale, or by analytically collapsing categories comprises the substance of the "disordered categories" controversy [338], [339]. In accord with the principle of taking up likelihood-based natural laws and not deductively making do with incomplete information, attention is focused here on the need to establish and maintain substantive commensurability in a shared frame of reference.

In finding equal heights in likelihood distribution maximums, and then tracing the line of coexistence between two solutions on the basis of a stochastic potential, Prigogine and Allen [188] (p. 27) are effectively modeling different categories' response likelihoods so as to link them and make them comparable. This is akin to addressing the problem of instrument commensurability in a form analogous to common sample or common item equating, where the meaning of what counts is connected across contexts [348] - [350]. They are emphatically not taking a phenomenological and empirical approach to statistical modeling in the manner of equipercentile equating, where "sampling fluctuation…is the subject of concern" [351] (p. 165). Here, numeric statistics that may be devoid of meaning are made comparable only in terms of their relative proportions.

Summarizing, both measurement theory's focus on stochastic invariance and Prigogine's sense of deterministic chaos can be seen as:

- distinguishing deductive population statistics motivated by sampling problems from inductive individual level observations motivated by the response process;
- contrasting likelihood and deterministic step functions describing an innovation's potential for surviving a varying number of generations; and
- modeling evolutionary sequences of biological or sociocognitive species inhabiting the same or varying environmental niches.

Measurement theory and the theory of dissipative structures differ in the rigor with which they conceive and implement the distinction between sampling probabilities and response likelihoods. Prigogine and his followers tend to remain methodologically locked into a statistical, analytic perspective. That is, with no concept of the need for and possibility of calibrating instruments embodying the laboratory model in real world applications, the idea of enacting the research results on broad scales in ways capable of mediating theoretical ideals and unique local circumstances is never brought up. Attention turns now to what this might entail.

IV. PRIGOGINE'S "TRINOMIAL CONJECTURE," SEMIOTICS, AND MEASUREMENT IN PRACTICE

De Castro and McShea [25] (p. 3), referencing what is here shown as Figure 6, say that, in positing a "trinomial conjecture," Prigogine asserted that:

> *…all complex dissipative structures, ranging from the classic 'Bénard instability' ... to organisms and human societies, [share a] common basis in the fluctuation–function–structure trinomial (Fig. 1). Prigogine thought of the trinomial as a bridge uniting the sciences, uniting the physical with the biological, social, and human.*

With no mention of Prigogine's trinomial, Bailey's [3] (pp. 20-28) social entropy theory points out the importance of adding a mediating third indicator term in two-level sociological models referencing only conceptual and empirical contrasts (Figure 7). In accord with Prigogine, however, Bailey explains that the levels are discontinuous and cannot be reduced in the manner of a numeric hierarchy, where higher level aggregates are the sums of the lower-level parts.

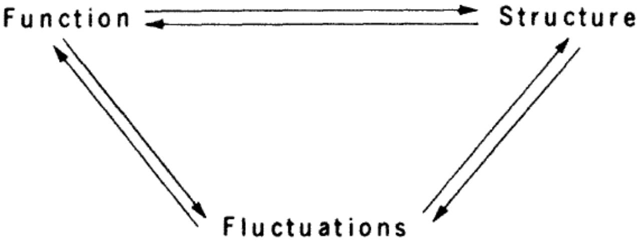


Figure 6. The "trinomial conjecture" (unnumbered figure from [16] (p. 120), [352] (p. 781)

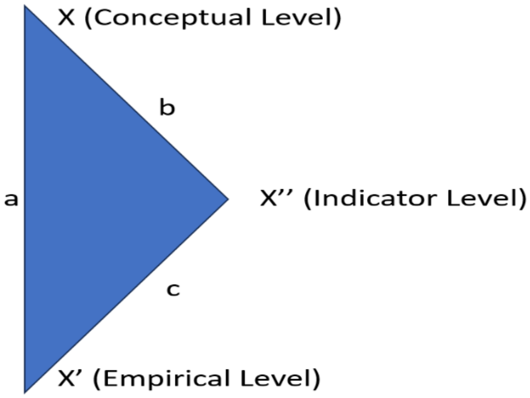


Figure 7. Social entropy theory variation on semiotic triangle, adapted from [3] (p. 26, Figure 2.1).

Bailey's semiotic distinctions between ideas, words, and things effectively generalizes the three terms in Prigogine's trinomial, where, within open systems, physical structures perform functions that may fluctuate, varying in ways that may dissipate entropy in the direction of increasing order and complexity. Bailey points out that, linguistically, structure is provided by conceptual ideals; function is embodied in semantic, alphabetic, and phonemic indicators; and fluctuations occur in relation to the empirical things represented. Brooks and Wiley [11] extending the semiotic analogy, see biological systems as self-referential, where structure corresponds with the genotype, function with the phenotype, and fluctuations with mutability [353]. In nature, each species forms a system integrating structural genotypes, functional phenotypes, and fluctuating mutability. In human affairs, analogously, each social form of life forms a system integrating formal conceptual structures, abstract media, and concrete local circumstances.

Further extending the analogy, we can see that, in human psychology, the superego structures conscientiousness and identity, while the id functions to embody subjective understanding, and the ego manages fluctuating daily events with varying effectiveness. In governance, constitutional law, rights, and the judiciary correspond with structure, while legislative measures and proportionate representation correspond with function, and executive administration, with mutable fluctuations. In science, theory and axioms structure formal laws, metrological unit standards and calibrated instruments function as abstract media, and experimental data fluctuate to reveal anomalous opportunities for new learning.

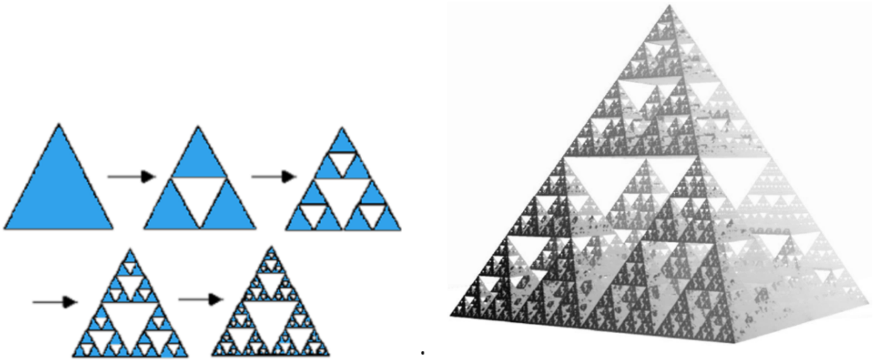


Figure 8. Sierpinski triangles and pyramid [355]

These analogical correspondences with the semantic triangle bring out the general relevance of semiotics as a model spanning the full range of fields, from physics to psychology [122] - [129], [353]. The fractally repeating pattern of Prigogine's trinomial across domains is aptly illustrated by the Sierpinski triangle or pyramid (Figure 8), a geometrical gasket that retains its properties of infinite area or volume at any level of magnification. This property of mathematically proportionate infinite area and depth is referred to as "scale-free" for reasons of its invariance but needs to be distinguished from the properties of measurement models capable of structuring consistent inferences across instruments, which also sometimes are referred to as "scale-free" [354]. The real-world relations of each set of triples across domains will surely distort the idealized images shown in Figure 8, but the illustration nonetheless aids in intuitively grasping the chaotic depth and breadth of the interpenetrating interdependent orders involved.

Dating from the 1867-1914 contributions of Peirce [356], and more recent efforts by Sebeok [129], [357], [358], a transdisciplinary semiotic theory of self-organizing systems integrates objective factuality and subjective meaning in a nondualistic worldview and approach to science [122] - [128], [353]. The expansion of semiotics from its original linguistic focus to broader applicability as a theory of evolving life in biosemiotics and ecosemiotics, and of information in cybersemiotics, has led to recognition of the need for its systematic operationalization in distributed cognitive ecologies' conceptual infrastructures:

> *A combination of cybernetic, systemic and semiotic understandings of the semiotics of information, cognition and communication area seems therefore crucial to the development of a systemic Cybersemiotics that can support teaching and human development* [124] (p. 20).

Philosophical investigations of meaning in science and measurement have developed corresponding theories of how spoken and written language serve as vehicles of thought [72], [81], [82], [120], [121], [130], [131], [135] - [137], [142], [359] - [362], and these are applied in psychology and social science as means of implementing nonreductive methods [108], [109], [311], [363] - [368]. Developmental psychology [87], [88], [90], [91], [95], [96], [369] has, in addition, independently developed self-referential theories of meaning and qualia toward analogous ends.

These initially separate trends have converged in research showing that stable psychological states and unstable transitions between them exhibit stage-level developmental structures, where measurement scaling research shows the existence of consistently separated discontinuous ranges [96], [156], [172], [246], [248] - [250], [370]. Fluctuations in this context occur in the process by which conceptual and/or behavioral integrations coordinate and align previously unarticulated background assumptions, making them objects of operations. Educators have then come to focus on facilitating learning as benefiting from environments in which playful experimentation is supported, misconceptions are elucidated by regular, actionable feedback, and failure is recast as the best way to improve outcomes [95], [97], [99], [156]. More will be said on this below.

Falmagne and Doignon [371] (p. 135) provide an explanation of the evolution of rationality from naivete to sophistication, formalized by a stochastic process with three interlinked parts that correspond with Prigogine's trinomial conjecture:

- Fluctuations: "One [part] is a Poisson process governing the times t1; t2; …; tn; ... of occurrence of quantum events of information, called tokens, which are delivered by the medium."
- Function: "The second [part] is a probability distribution on the collection of all possible tokens, which regulates the nature of the quantum event occurring at time n. "
    - "Any token is formalized by some pair xy of distinct alternatives, bearing a positive or negative tag. "
    - "The occurrence of a positive token xy signals a quantum superiority of x over y, while the corresponding negative token y indicates the absence of such a superiority."

- Structure: “The last part of the stochastic process is a Markov process describing the changes of states occurring in the subject as a result of the occurrence of particular tokens.”

Similar ways of conceptualizing shifts across levels of complexity are described in the measurement perspectives presented in Wilson’s [248], [250], [372] Saltus model, Commons’ [373] model of hierarchical complexity, and K. Fischer’s [94], [95] skill theory. In each case, concrete fluctuations in empirical experience coalesce into repeating abstract patterns akin to the schooling, swarming, or flocking behaviors of fish, insects, or birds, as illustrated in Ireland and Statsenko [374] (p. 96, Figure 2). The coherence of these patterns induces analogical and metaphorical associations at the formal, structural level of complexity when new experiences can be related to previous ones. Gaining a clue, hint, or hunch of a salient source of value to anticipate and look for in the world makes it possible for a descending dialectic to proceed by naming the abstract pattern via a metaphor, sharing it in communications, and applying it in new, previously unmet instances of the experience to negotiate shared understandings with others.

V. INFORMATIONAL ENTROPY VS THERMODYNAMIC ENTROPY IN MEASUREMENT

There are enormous variations in the explicit and implicit meanings and mathematics associated with both entropy and information, so many, in fact, that sorting them all out seems an interminable task [29] (pp. 9-11). Though it will be impossible to fully achieve, I would like here to try to not contribute to additional confusion. To begin, although the definition of entropy as a fixed amount of information in communications theory since Shannon is formally analogous to Boltzmann’s thermodynamic entropy function, there are important reasons for recognizing and actively distinguishing between them. Shannon and Weaver defined the measurement of information in terms of a discrete probability distribution P where the entropy H of P is given in a log base two function termed bits. The entropy H of P attains its maximum value of log base 2 of the sample size n when P is equally probable across the entire range of the distribution. The quantity of information needed to decide between two equally probable possibilities is termed a bit, and the number of bits required to obtain a complete statistical description of a system determines its overall information content.

In Shannon’s [375] (pp. 50-57) sense, a message’s information content, then, is equal to its entropy in the sense that the most informative string of symbols of a given length will be random, where each symbol is equally probable. Redundancy and exceptions, in contrast, create divergent differences in symbol probabilities, as each is no longer equally unique. Shannon information content and entropy are then reducible in ways that are not always interpretable physically [11] (p. 65), leading to Haynes et al.’s [1] suggestion that mature applications of entropy in psychology and the social sciences should eschew connections with thermodynamics.

An opposite definition of the relation of entropy and information is given by Layzer [376], who sees them as inversely constrained by a simple conservation law holding that their sum is constant and must be equal to the maximum possible in a given situation. Now, in Layzer’s sense, entropy decreases as complexity and information increase, so that evolution becomes aligned with the latter in the context of evolving dissipative structures. This does not, however, change the overall characterization of the statistical sense of probability employed. Shannon’s statistical sampling perspective on the content internal to messages defines maximum information content mechanically, so the whole is the sum of the parts. There is no distinction between levels of complexity.

No mention is made of semantic closure [353] or of the receiver’s semantic reaction [377], [378], which involve the capacity for the content to inform meaningful inferences, interpretations, or actions outside of the communications channel. Though Weaver [379] (pp. 25-28) raises these issues, he minimizes their importance, saying they likely imply only “minor additions, and no real revision” (p. 26). Weaver unabashedly accepts that “the concept of information developed in this theory…has nothing to do with meaning, and…deals not with a single message but rather with the statistical character of a whole ensemble of messages” (p. 27). The focus on information content is then entirely external to the communication. Unexpected intrusions of surprising and improbable symbols, or of emphatic repetitions of urgently important symbols, reduce information and entropy in a mechanically dynamic way in this statistical perspective even though they may represent essential features of the environment demanding close attention. Information so defined is, however, incapable of representing itself recursively, and so sets up closed systems incapable of evolving.

A measurement modeling perspective on recursive processes [246], [249], [250], [369], and on fractal, chaotic processes [380], [381], in contrast, focuses on irreducible variation in the contributions of each symbol and each message to an overall composite meaning that is greater than the sum of the parts. Now, information is defined in terms of the sufficiency of consistent, repeatable, monotonic likelihood functions. These likelihoods stochastically fluctuate within semantically closed loops structured in the flows of open systems, just as matter and energy do.

VI. PRACTICAL IMPLICATIONS OF ENTROPY-DISSIPATING MULTI-LEVEL COMPLEXITY

More specific points of contact between the nonequilibrium thermodynamics of entropy-dissipating processes and social processes then involve ways in which Prigogine’s sense of a new dialogue of humanity with nature implies:

- the objective repeatability of structures Prigogine describes in terms of self-organized “deterministic chaos” and which measurement theory postulates in terms of reproducible patterns of stochastic invariance that demonstrably maintain their properties across samples, observers, and instruments [348], [382] - [390];
- the shift Prigogine sees away from a sense of probabilistic models connoting a resigned acceptance of incompleteness and ignorance toward probabilities that are foundational expressions of scientific laws, which aligns with measurement theory’s perspective on differences between deductive population-level probabilistic models motivated by sampling issues and inductive individual-level likelihood models motivated by the response processes generating the data;
- physical processes occurring at the quantum level that Prigogine and others suggest inform an objective ontology that removes the need for an observer to play an “extravagant” role in the evolution of nature, and which accords with developments in the domain of semiotics concerning the ways in which physical, biological, and information processes can be seen to read, write, and measure themselves; and
- violations of the traditional binary logic of the law of the excluded middle and the law of noncontradiction, which resonate not only with measurement theory but also with parallels in the multivalued, nonternary logics posited in aesthetics, ethics, existentialism, hermeneutic phenomenology, and feminism; further research in this regard is in process.

The implications of this new dialogue with nature for new dialogues among humans suggest a pragmatic and actionable program for advancing political economies explicitly intended to take the form of participatory social ecologies. Such ecologies, like those in nature, involve micro, meso, and macro forms of life exhibiting within-individual, individual, and population nonequilibrium processes. These levels of complexity are currently inconsistently recognized in widely institutionalized information systems that sometimes rightly distinguish between, for instance, numeric counts of rocks and measured quantities of rock, while also not distinguishing between numeric counts of correct answers or points of agreement, and measured quantities of abilities or attitudes [391] - [393].

Though current organizational systems persistently confuse and fail to separate these levels into distinct but interrelated domains, calls for consistently achievable models and methods for doing so date back several decades [394] - [399]. Hayman et al. [396] (p. 31), for instance, say:

> *Our hypothesis is that: The utility of a set of evaluation data varies inversely with the number of organizational levels between the action the data describe and the decision process they are intended to influence.*
>
> *In other words, the closer a set of data is to the organizational level for which it will be used (for decision-making), the more useful the data will be.... Conversely, the principle states that the further a set of data is from the organizational level for which it will be used (for decision-making), the less useful that data will be.*

According to this hypothesis, then, because of the great distances separating fragmented and incommensurate local data systems from the global decision processes they are intended to influence, efforts like the United Nations Sustainable Development Goals and the Carbon Disclosure Project must inevitably fail. Worldwide science and commerce conducted in the mathematical language of the SI units has succeeded on huge scales, altering its environment and thereby compromising its potential for continued productivity. Global systems are causing human suffering, social discontent, and environmental degradation that will be successfully addressed only when approached at the relevant organizational level. These problems will be made tractable only when, in the manner of a Chinese finger puzzle, humanity ceases applying the reductionist and dualistic top-down, outside-in methods causing the problems and relaxes into the nondualism of a new class of SI units embodying metrological solutions distributing irreducibly complex unified subject-object technologies globally [110], [400] - [404]. Prigogine’s thesis of a “new intelligibility” and a new science of collective rationality provides conceptual tools and perspectives essential to meeting 21st-century challenges at a level offering viable possibilities for their solution.

Rousseau [399] also takes up the ecological fallacy and the organizational problems that follow when individual relationships are over-generalized, and issues of level are not considered in design as well as in data aggregation. In Star and Ruhleder’s [405] (p. 118) terms:

> *If we, in large-scale information systems implementation, design messaging systems blind to the discontinuous nature of the different levels of context, we end up with organizations which are split and confused, systems which are unused or circumvented, and a set of circumstances of our own creation which more deeply impress disparities on the organizational landscape.*

Ireland and Statsenko [374] expand on the theme by providing extensive breakdowns of the organizational implications of complex systems theory. Though they do so in the usual manner that assumes centralized planning and data analysis to be the only available options for policy and programs, a point of departure toward metrological models of distributed cognitive ecosystems is suggested by their recognition of the fractal nature of project management and the relevance of mathematical power laws [374] (p. 95). They illustrate the fractal parsing of project responsibilities and processes using a Sierpinski triangle, though it is not named as such, and they helpfully state the problems discontinuous demands on modeling impose by having to address both the entire organization’s needs as well as individual decisions and behaviors.

Multilevel log-interval measurement models are in fact instantiations of the kinds of power laws referenced by Ireland and Statsenko. Their applicability to problems of organizational complexity are further suggested by Ostrom [406] (p. 50), who argues that:

> *The rules affecting operational choice are made within a set of collective-choice rules that are themselves made within a set of constitutional-choice rules. The constitutional-choice rules for a micro-setting are affected by collective-choice and constitutional-choice rules for larger jurisdictions.*

Scott [40] shows that the history of humanity's autocrats and despots demonstrates just how far reductionist higher order constitutional- and collective-choice rules can be invented and imposed to serve the interests of an elite at the expense of the greater good. He [40] (pp. 355-357) suggests, without mentioning semiotics by name, that an alternative nonreductionist approach could be formulated by taking language as a model, so that the emergent, self-organized sociohistorical processes by which meaningful things are brought into words comprise the basis of policies and programs.

That kind of a possibility for operationalizing Ostrom's distinctions between organizational levels is then apparent in Hayek's [116] (p. 88), [117] goal of enabling people to do the desirable thing without having to be told. Meaningful freedom of choice constrained within the limits of a sociocognitive ecological niche might be hypothesized to follow from operational choices induced by naturally cohering, self-organized collective choice rules, which in turn are induced by constitutional choice rules that can similarly be shown to assert an independent, repeatable and reproducible existence.

A parallel process can be seen in Piaget's [86] emphasis on the importance of play in learning and development as how children's capacities are induced in relation to the learning environment across preoperational, concrete, and abstract levels of complexity. Nersessian's [92] sense of scientific experimentation as play works toward a common end and substantiates Dewey's [407] (p. 201) characterization of schools as laboratories for experimenting with life. Fleshing out these parallels, we can see that, in Gadamer's [89] (pp. 104, 367), [366] terms, answers to questions in dialogue are operational choices playfully induced by collective choice rules represented in the available grammatical and phonemic standards; these latter are themselves induced by the conceptual determinations structuring them as constitutional choice rules.

Integrating R. Fisher's [189], [190] and Rasch's [63], [65], [210] terms for Ostrom's level distinctions, operational choices are response likelihoods induced by the collectively coherent differences between people's, objects', or processes' abilities or capacities and the difficulties or obstacles they confront. The collectively structured rationality of the construct measured—a learning progression, developmental sequence, or healing trajectory—is then induced by constitutional choice rules demonstrably established by explanatory models [43], [44], [250], [255], [265], [326], [369], [408] - [419] predicting item calibrations and person measurements from their characteristics.

Educational, health care, professional certification, and other systems integrating all three of these levels to various degrees have existed for decades [234], [256], [420], [421], but metasystematic and paradigmatic integrations [52], [53], [254], [255], [258], [283], [422] have scarcely begun.

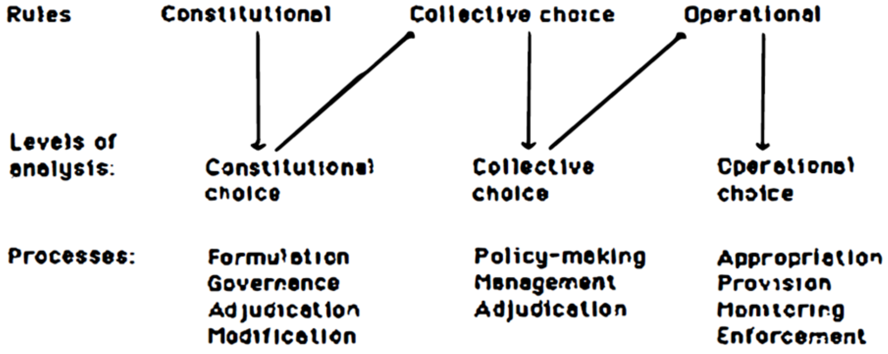


Figure 9. Linkages among rules and levels of analysis [397] (Figure 2.2, p. 52)

Figures 9 and 10, from Kiser and Ostrom [397], [406] illustrate linkages across levels of analysis, where operational choice rules are embedded within collective choice rules, and both are situated in the context of constitutional choice rules. The challenge is to design data, information, and knowledge systems with close attention to the distinct analytic and reporting needs of each level vis a vis the creation of common languages capable of embodying and mediating:

- constitutional choice rules involving the formulation and modification of conceptual ideals explaining and predicting governance and adjudication decisions and behaviors;
- collective choice rules informing standards for policy-making and managerial adjudications; and
- operational choice rules addressing empirical appropriations, provisions, monitoring, and enforcement.

Semiotically modeled decisions systematically integrate probabilistic forms of knowledge across an irreducible array of discontinuous levels of complexity. They can do so in the context of probabilistic laws of human nature by attending to the interdependencies of the three levels of complexity [423]:

- constitutionally via conceptual maps of measured constructs aiding in the design of measuring instruments justified by explanatory theory demonstrating in a publicly reproducible way a transparent understanding of the object of the conversation [424], [425];
- collectively via standardized Wright maps of the locations of persons and items on the quantitative continuum expressed in the medium of a publicly available common language and metric read from a calibrated instrument, with associated uncertainty and data quality statements [235], [415]; and
- operationally via response maps of the empirical evidence of individual responses useful in guiding instruction, clinical care, management, etc. [426] - [428].

Figure 10 illustrates the involvement of both formal and informal collective choice arenas in operational rules in use. In well-designed systems, (a) local operational decisions concerning the unique situations of individuals will be afforded opportunities for adaptive accommodations lacking in less well-designed systems, with clearly stated tolerances for uncertainties; and (b) formal collective decisions concerning accountable regulatory matters will incorporate uncertainty estimates useful in determining the probabilities of distinct problems in conformity or compliance [56].

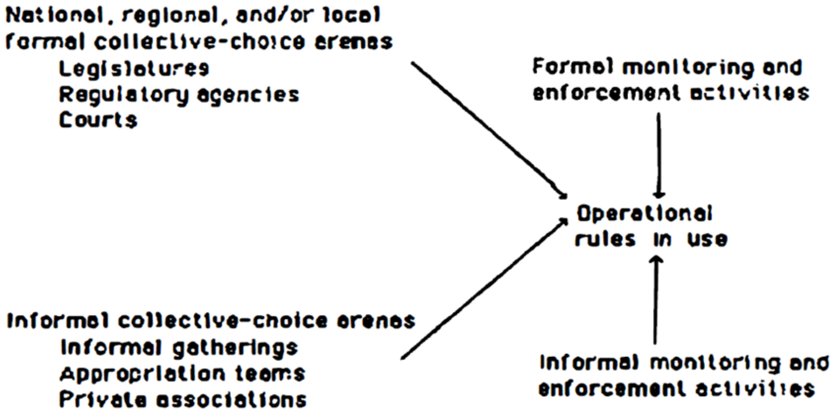


Figure 10. Relationships of formal and informal collective-choice arenas and common pool resource operational rules [397] (Figure 2.3, p. 53)

In this context, in contrast to the compulsory red tape of bureaucracies focused on rigid compliance with sometimes irrelevant policies and procedures, information quality can be fit to the needs of the moment. This can be accomplished by supporting the decision process with information sufficient to the task, which need be neither overly precise or imprecise. Explicitly incorporating uncertainty into the foundations of the definition of knowledge demands estimating it as a factor impacting decisions and behaviors in formal, high-stakes situations. When a fit-for-purpose degree of uncertainty supports the assertion of a decision, that outcome ought to be made scientifically, legally, and financially defensible and accountable, so far as possible.

The importance of distinguishing levels of analysis is plainly key to the reasons why Ostrom's and Star's ideas have been combined in work on infrastructuring participatory social ecologies [429] - [431], though few or no overt references are made in that literature to developmental levels of hierarchical complexity or to metrologically relevant measurement modeling.

VII. CONCLUSIONS

Measurement systems that incorporate "the collective dimension of individual actions" as prescribed by Prigogine and Allen [188] (p. 37) do so by modeling individual responses projecting stochastic invariances at higher orders of complexity. Fluctuations in response likelihood may cohere and exhibit instabilities leading to the emergence of self-organized orderly patterns. These patterns are inaccessible to what Prigogine calls the "collective utility function" of a deterministic expectation of simple static systems. They have, however, long informed practical applications in the meaningful structuring of measurements reliably connecting persistently reproducible invariances with defined metrics in common languages addressing different communities of research and practice in their own terms at their own operational, collective, and constitutional levels of complexity.

Prigogine's proposal that all far-from-equilibrium processes are characterized by evolutionary capabilities realized by the interactions of structures, functions, and fluctuations is supported by the identification of analogous three-part conceptualizations of interacting levels of complexity in multiple other fields. Prigogine's mathematical formulation of probabilistically structured natural laws distinct from deductive statistical descriptions of population dynamics enables a principled engagement with his expectation that his theory of dissipative structures should be relevant to Piaget's theory of cognitive development. The extension of Piaget's account of developmental levels in the model of hierarchical complexity fleshes out Prigogine's theory to the extent that the mathematical terms of the models informing log-interval scales for the measurement of cognitive development generalize the fractal repetition of the evolving structures.

One further emphatic stress on a key consequence of irreducible complexity must be made, given the intention of here of provoking a clear grasp of the implications of a shift from a dualistic and reductionistic conception of the intelligence of isolated individuals toward a nondualistic collective rationality projected at a higher order level. What is emerging from these considerations is a rearticulation of subjective experience, an expansion on the philosophical critique of subjectivity that has been underway for centuries [432], [433]. The intention to complement today's predominant orientation toward a deductive sense of sampling probabilities with a new emphasis on inductive response likelihoods requires a shift away from a conception of subjectivity as alienated from and adapting to an independent objective reality toward a sense of subjectivity as being objectively induced by its constitutive interrelationships with its environment.

A basic manifestation of that interrelationship infuses a fundamental assumption permeating many conceptions of logic and rationality: that explicit conscious control of cognitive operations is a hallmark of science and reason. The force of the cultural environment as a factor shaping the content and limits of individual thinking is truly profound, ironically extending even to the point of making individual's isolated mental operations seem to comprise the essence of cognition. Bateson [102] (p. 145) reports that,

> *Freud, even, is said to have said, 'Where id was, there ego shall be,' as though such an increase in conscious knowledge and control would be both possible and, of course, an improvement. This view is the product of an almost totally distorted epistemology and a totally distorted view of what sort of thing a man, or any other organism, is.*

Diametrically opposed to this view is the common experience of learning how to perform a wide variety of physical and cognitive processes so well that they become intuitive. The "automaticity" of this kind of learning is a well-recognized goal in education [434], [435]. Learning to execute a triple axel, to dance the frug, to pitch a strike, to play the cello, or to wash the dishes do not involve conscious calculations of force, mass, and acceleration that are then mechanically applied to move the body. As Samuel Butler put it, "...the better an organism 'knows' something, the less conscious it becomes of its knowledge" [102] (p. 138).

Contrary, then, to Kahneman's [436] urge for us to think more slowly and carefully, to Rose's [437] and Nadler and Shapiro's [438] strategies for replacing bad thinking with good, and to Stengers' [439] call for "slow science," stands Hayek's [440] (p. 88} concurrence with Whitehead's [441] (p. 61) observation that:

> *It is a profoundly erroneous truism, repeated by all copy-books and by eminent people when they are making speeches, that we should cultivate the habit of thinking of what we are doing. The precise opposite is the case. Civilization advances by extending the number of important operations which we can perform without thinking about them.*

Languages and technologies facilitate the performance of operations we do not understand—such as when reading a thermometer requires no grasp of thermodynamics—by providing a medium objectively extending—inducing—the subjective, embodied experience of a phenomenon [74], [78], [442], [443]. In the same way that a blind person extends the perceptual reach of their senses by using a cane, so also do overlearned musical scales allow jazz artists to collaboratively create intuited improvisations [444]. Everyday language is also overlearned to the point of enabling fluency: the automatic, involuntary associations that flow so quickly that it is virtually impossible to apprehend them before they are made. Oral language is learned physically in terms of the resonances of sounds in the body and the rhythmic inhalations and exhalations of air; we skip over too quickly the importance of the embodied intuitions developed in language use as they provide important clues as to how understanding develops relationally [445].

Just as children form new concepts via complex interactions in physical experience that lead to capacities for the "fast thinking" Kahneman [436] would like to slow down, so also are similar processes systematically cultivated in science. As Prigogine noted, thermodynamics, after all, "was formulated not by the theoreticians of mechanics who contemplate the world, but by the engineers and physical chemists who deal with the world through their experiments" [17] (p. 245).

This perceptive observation is supported by the historical evidence, and may have been prompted by Prigogine's colleague, Stengers, a noted philosopher of science [42], [439], [446]. Kuhn [222] (p. 90), for instance, points out that, of the nine researchers contributing to the quantification of energy conversion processes, seven "were either trained as engineers or were working directly on engines." As was also described by Nersessian [92], the importance of working in close proximity with technologies enabling repeatable experiences in reproducing and exploring controlled effects led de Solla Price [447] (p. 240) to similarly remark that "thermodynamics owes far more to the steam engine than ever the steam engine owed to thermodynamics," and that "historically the arrow of causality is largely from the technology to the science." This historical and ontological priority of technology over science motivates the emergence of the concept of technoscience taken up in science and technology studies

[81], [106], [137], [448], [449]. But the effect of success in advancing the science has been to make the objectively repeatable automatic associations methodically built into tools, standards, textbooks, and professional associations into media for virally communicated social contagions in which the map is mistaken for the territory and the individual mental model for the shared collective rationality.

Thinking, then, as a process of conscious, creative, original, reasoned deliberation is quite rare, because rationality is bound and defined by the limits of the automatic associations facilitated by language [450]. What Kahneman [451] (p. 1450) calls "effortless associative thinking" in bounded rationality corresponds with

- Gadamer's [89] (p. 463) assertion that it is truer to say language speaks us than vice versa;
- Wittgenstein's [452] (p. 74), [453] recognition that "the limits of my language mean the limits of my world;" and with
- Mach's [454] (pp. 481-495) [455] - [457] sense of an economy of thought in which the shared symbol system lowers communications' transaction costs.

Distinguishing the poetics of creative invention from methodical reproductions of proven results, Heidegger [458] (p. 135) accordingly notes that "science itself does not think," meaning that the routine practice of laboratory science takes place in terms that certainly differ across experimental, metrological, and theoretical communities of practice, but which accept the need to operate within well-defined conceptual borders. This state of affairs accounts for Latour's [106] (pp. 249-250) assertion that growing new sciences entail expanding metrological networks; as he [71] (p. 210) put it, "Through the materiality of the language tools, words finally carry worlds."

The same principle applies in the way economic expansions often entail the lower transaction costs afforded by currency unions and metric standards [459]. And Wittgenstein [460] (p. 47) then defines philosophy as a constant battle against bewitchment by language. But instead of struggling against the subjective play of language games contingent on the existing semiotic complex of ideas, words, and things, should not it be possible to imagine other semiotic complexes in which the games played are less viciously circular, self-destructive, and self-defeating?

The opportunities opened up by the implications for practice explored in this essay suggest that there are ample reasons justifying the reconception of the problem as one of learning to free ourselves by relaxing into the finger puzzle trap. The low transaction costs afforded by the availability of premade idea-word-thing associations create an economy of thought from which it is incredibly difficult to leave. New ideas expressed in new words in relation to unfamiliar things rarely are amplified into wide circulation. To do so, they must represent a clear adaptive advantage—perhaps to an environment inducing unsustainable strategies—offering greater economic, social, or other efficiencies, efficiencies which must be broadly understood in not just logical terms but in terms of emotions, politics, aesthetics, and ethics, as well. In contrast to the prevailing view that humans rationally evaluate situations so as to act in their own self-interest, what actually happens is better described in terms of bounded rationality, where behaviors are imitated and analogies are propagated through social networks structured to facilitate virally communicable contagions of meaning and care, many of which may sacrifice long term survivability in favor of getting through the day [13], [31], [461] - [463].

To say the least, a large array of further questions stands in need of further research. Is a nonreductionist model of the structure, functions, and fluctuations of evolutionary change across physical, chemical, biological, psychological, social, and economic domains possible? Can a science of entropically-driven evolving processes occurring far from thermodynamic equilibrium be conceived and implemented at a cross-paradigmatic level of complexity? Such a science would have to be able to supplant the mechanically conceived theory, instruments, and data of the contemporary Western, Cartesian, dualistic paradigm with those of an emerging global nondualistic paradigm. Each domain would have to be energized by the physical work performed by dissipated entropy production in each sphere's interacting structures, functions, and fluctuations. Order-of-magnitude shifts in the efficiencies of entropy dissipation would be induced by the infrastructure of each newly emergent level of complexity integrating the previous level's operations with enhanced dissipative opportunities afforded by the external environment.

In accord with that model, in the emerging paradigm, metasystematic paradoxes of heterogeneously unified and disunified, harmonious and dissonant, voluntary and involuntary individual operations would have to be integrated into shared symbol systems powering language as the vehicle of thought. The kinds of questions emerging here include but are not limited to the following:

- Might we finally be on the verge of productively fulfilling Bohr's expectation that, having encountered a paradox, there is some hope of progress [464] (p. 140), [301] (p. 686)?
- Might R. Fisher's [189], [190] distinction between deductive population-level probabilities and inductive individual-level likelihoods, and Rasch's incorporation of the structure of natural law in his models of individual-level measurements, be combined to create an entropy-based SI unit defining a scaling factor analogue of Boltzmann's constant in information theory, which today remains just as unavailable as it was in 1959, when von Foerster [36] (p. 19) lamented its absence?
- Restating that question, can shared languages and common metrics calibrated as representations of simultaneously global generalities and local specificities be operationalized as the nonequilibrium

thermodynamic engines energizing labor-saving processes and lifting the burden of initiation in an economy of thought?
- Might the dissipation of entropy inform the coherence of a new class of SI units in a manner analogous to the way the speed of light does for the existing SI?
- In other words, how might Prigogine's sense of a new intelligibility inform or contrast with the new science of enchantment anticipated by Sahlins [465], or the evolving participatory complexity described by many others [30], [33], [301] - [304], [436], [466]?
- Will humanity find ways of operationalizing nonWestern and premodern traditions in a new paradigm that relinquishes dualistic reductions to isolated individuals in favor of cybernetic feedback loops in mutually causal, interdependent relationships?
- Can the similarity of the logical-mathematical structures common to nonequilibrium processes across physics, biology, and psychology inform a new paradigm transforming the thermodynamic "metrisation of the economic state space" begun by von Neumann, Leontief, and others [9], [467] - [469]?
- Can we imagine how models of collective rationality might be embedded in the rules, roles, and responsibilities of a new institutional economics of human, social, and natural capital?
- Can humanity learn to think globally together at the same time it acts locally as uniquely situated individuals?
- Can the seemingly irresistible cultural power of the dominant paradigm's categorical reductions be circumvented in ways that enable the organic cultivation of a new paradigm of irreducible complexity?
- Might this be facilitated by shifting the theory of games in economics away from centrally planned analytics and assumptions of hyperrational consumers toward a more authentic sense of play truer to the distributed participation of players whose undivided attention is caught up in the flow of events?
- Can the inherent insufficiency of systematically fragmented solutions applied to global problems be recognized and addressed in time to prevent clearly impending climate disasters?
- Might individuals be empowered to profitably innovate new social technologies on mass scales by new institutional economies, successfully addressing global problems at the global scale where they exist?
- Will the long-sought goal of reducing all sciences to physics be accomplished, ironically, not only by means of irreducible complexity but also by constituting the identity of each science as an independent domain of investigation?
- Might governance become an instrumentally mediated integration of art and science, like music?

Questions in this vein take up the general applicability of a semiotic model of hierarchical complexity as a matter in need of close attention. Could a coherent system of entropy-based metrological unit standards feasibly structure communications and lower transaction costs in a multilevel ecological economy of human, social, and natural capital markets? The challenges humanity faces in moving toward needed solutions are immense; but having formulated the problem, human ingenuity may yet win the day.

ACKNOWLEDGMENT

The research reported here was not supported by external funding. The author acknowledges the kind moral support and ongoing conversations of multiple colleagues over many years and accepts all responsibility for any errors and omissions.

[1] Haynes, K. E., Phillips, F., Yu, M. C. Y. Recent applications of entropy in social and environmental sciences. In *Handbook on Entropy, Complexity and Spatial Dynamics;* Reggiani, A., Schintler, L. A., Czamanski, D., Patuelli, R., Eds.; Edward Elgar Publishing, Northampton, Massachusetts, USA, 2021; pp. 32-45.
[2] Georgescu-Roegen, N. *The entropy law and the economic process.* Harvard University Press: Cambridge, MA, USA, 1971.
[3] Bailey, K. D. *Social entropy theory*. State University of New York Press: Albany, NY, USA, 1990.
[4] Binswanger, M. From microscopic to macroscopic theories: Entropic aspects of ecological and economic processes. *Ecological Economics*, **1993**, *8*, 209-233.
[5] Floyd, J. Thermodynamics, entropy and disorder in futures studies. *Futures*, **2007**, *39*, 1029-1044.
[6] Mayumi, K.; Giampietro, M. Entropy in ecological economics. In *Modeling in ecological economics*. Edward Elgar, Cheltenham UK, 2004, 80-101.
[7] Zencey, E. Some brief speculations on the popularity of entropy as metaphor. *Metaphor and Symbol*, **1991**, *6*, 47-56.
[8] Gyarmati, I. *Non-equilibrium thermodynamics*. Springer, Berlin, 1970.
[9] Brody, A. Physical (phenomenological) economics? (A semi-centennial for J. von Neumann and W. Leontief). *Acta Oeconomica*, **1989**, *41*, 257-266.
[10] Ackerman, F. Still dead after all these years: Interpreting the failure of general equilibrium theory. *Journal of Economic Methodology*, **2002**, *9*, 119-139.

[11] Brooks, D.R.; Wiley, E.O. *Evolution as entropy: Toward a unified theory of biology, 2nd ed.* University of Chicago Press, Chicago, Illinois, USA, 1988.

[12] Brooks, D.R.; Wiley, E.O.; Collier, J.D.; Maurer, B.A.; Smith, J.D.H.; E. O. Willey, E.O. Entropy and information in evolving biological systems. *Biology and Philosophy*, **1989**, *4*, 407-432.

[13] Platt, J.R. Social chain reactions. *Bulletin of the Atomic Scientists: Man and his habitat, Part II*, **1961**, *17*, 365-386, available: New York. *The step to man*. John Wiley & Sons. 1966, 39-52.

[14] Prigogine, I. *Introduction to thermodynamics of irreversible processes, 2nd ed.* Wiley, New York, NY, USA 1961.

[15] Prigogine, I. Unity of physical laws and levels of description. In Prigogine, I.; Grene, M. eds. *Interpretations of life and mind: Essays around the problem of reduction*. Humanities Press, New York, NY, USA 1971, 1-13.

[16] Prigogine, I. Order through fluctuation: Self-organization and social system. In Jantsch, E.; Waddington, C. eds. *Consciousness and evolution: Human systems in transition*. Addison Wesley, New York, NY, USA 1976, 93-130.

[17] Prigogine, I. Physics and metaphysics. In Lawrence, J.H.; Gofman, J.W.; Hayes, T.L. Eds. *Advances in biological and medical physics, Vol. 16*. Academic Press, New York, NY, USA 1977, 241-263.

[18] Schieve, W.C.; Allen, P. Eds. *Self-organization and dissipative structures: Applications in the physical and social sciences*. University of Texas Press, Austin, Texas, 1982.

[19] Achermann, G.; De Luca, G.; Simoni, M. An information-theoretic and dissipative systems approach to the study of knowledge diffusion and emerging complexity in innovation systems. In *International conference on computational science*. Springer, Cham, Switzerland, 2020, 252-265.

[20] Andrade, E. Integration of thermodynamic, quantum and hierarchical theories of information in the context of Peircean semiosis--A review. *BioSystems*, **2014**, *120*, 10-20.

[21] Arthur, W.B. Complexity economics: A different framework for economic thought. In Arthur, W.B., ed. *Complexity and the economy*. Oxford University Press, New York, NY, USA 2014, 1-29.

[22] Artigiani, R. Revolution and evolution: Applying Prigogine's dissipative structures model. *Journal of Social and Biological Structures*, **1987**, *10*, 249-264.

[23] Bailey, K.D. Living systems theory and social entropy theory. *Systems Research and Behavioral Science*, **2006**, *23*, 291-300.

[24] Bailey, K.D. Boundary maintenance in living systems theory and social entropy theory. *Systems Research and Behavioral Science*, **2009**, *25*, 587-597.

[25] de Castro, C.; McShea, D.W. Applying the Prigogine view of dissipative systems to the major transitions in evolution. *Paleobiology*, **2022**, 1-18.

[26] Dinga, E. Tanasescu, Social entropy and normative network. *Entropy*, **2020**, *22*, 1051.

[27] Dixon, P.W. On Amit Goswami and Maggie Goswami's quantum creativity: Waking up to our creative potential: Entropy, information theory, and dissipative structures. *Creativity Research Journal*, **2002**, *14*, 439-440.

[28] Giampietro, M. Multi-scale integrated analysis of societal and ecosystem metabolism. In Villamayor-Tomas, S.; Muradian, R., Eds. *The Barcelona School of Ecological Economics and Political Ecology: A Companion in Honour of Joan Martinez-Alier*. Springer International Publishing, Cham, Switzerland, 2023, 109-121.

[29] Haddad, W.M. Thermodynamics: The unique universal science. *Entropy*, **2017**, *19*.

[30] Hidalgo, C. *Why information grows: The evolution of order, from atoms to economies*. Basic Books, New York, NY, USA 2015.

[31] Hodas, N.O.; Lerman, K. The simple rules of social contagion. *Scientific Reports*, **2014**, *4*, 4343.

[32] Laszlo, E. Technology and social change: An approach from nonequilibrium systems theory. *Technological Forecasting & Social Change*, **1986**, *29*, 271-283.

[33] Laszlo, E. Ed. *The new evolutionary paradigm: Keynote volume*. Routledge, London, UK, 2019.

[34] Mitar, M. Bailey's social entropy theory as an explicit theoretical approach for an empirical assessment of security of contemporary societies. *Quality and Quantity*, **2010**, *44*, 941-955.

[35] Mitchell, M. *Complexity: A guided tour*. Oxford University Press, New York, NY, USA 2009.

[36] von Foerster, H. On self-organizing systems and their environments. In *Understanding understanding: Essays on cybernetics and cognition*. Springer, New York, NY, USA 2003, 1-19.

[37] Prigogine, I.; Stengers, I. *La nouvelle alliance: Métamorphose de la science*. Gallimard.

[38] Prigogine, I.; Stengers, I. *Order out of chaos: Man's new dialogue with nature*. Bantam Books, New York, NY, USA 1984.

[39] Prigogine, I. A new rationality? *Mondes en développement*, **1986**, *14*, 19-39.

[40] Scott, J.C. *Seeing like a state: How certain schemes to improve the human condition have failed*. Yale University Press, New Haven, Connecticut, USA, 1998.

[41] Jasanoff, S. Technologies of humility: Citizen participation in governing science. *Minerva*, **2003**, *41*, 223-244.

[42] Stengers, I. The challenge of ontological politics. In de la Cadena, M.; Blaser, M., Eds. *A world of many worlds*. Duke University Press, Durham, North Carolina, USA, 2018, 83-111.

[43] Melin, J.; Cano, S.; Pendrill, L. The role of entropy in construct specification equations (CSE) to improve the validity of memory tests. *Entropy*, **2021**, *23*, 212.

[44] Melin, J.; Cano, S.; Flöel, A.; Göschel, L.; Pendrill, L. The role of entropy in construct specification equations (CSE) to improve the validity of memory tests: Extension to word lists. *Entropy*, **2022**, *24*, 934.

[45] Melin, J.; Pendrill, L.R. The role of construct specification equations (CSE) and entropy in the measurement of memory. In Fisher, W.P. Jr.; Cano, S.J., Eds. *Person-centered outcome metrology*. Springer, Cham, Switzerland, 2023.

[46] Massof, R.W.; Bradley, C. Constructing a latent disease state variable from all relevant clinical examination data. In Fisher, W.P. Jr.; Pendrill, L.R., Eds. *Models, measurement, and metrology extending the SI*. De Gruyter, Berlin, Germany, 2024.

[47] Pendrill, L.R.; Melin, J.; Cano, S.J.; EMPIR NeuroMET 15HLT04 consortium, Metrological references for health care based on entropy. *19th International Congress of Metrology (CIM2019)*, **2019**.

[48] Fisher, W.P. Jr.; Cano, S. Eds. *Person-centered outcome metrology: Principles and applications for high stakes decision making*. Springer, Cham, Switzerland, 2023.

[49] Fisher, W.P. Jr.; Massengill, P.J. *Explanatory models, unit standards, and personalized learning in educational measurement: Selected papers by A. Jackson Stenner*. Springer, Cham, Switzerland, 2023.

[50] Mari, L.; Wilson, M. An introduction to the Rasch measurement approach for metrologists. *Measurement*, **2014**, *51*, 315-327.

[51] Mari, L.; Wilson, M.; Maul, A. *Measurement across the sciences: Developing a shared concept system for measurement, 2nd ed.* Springer, Cham, Switzerland, 2023.

[52] Melin, J.; Cano, S.J.; Flöel, A.; Göschel, L.; Pendrill, L. Metrological advancements in cognitive measurement: A worked example with the NeuroMET memory metric providing more reliability and efficiency. *Measurement: Sensors*, **2023**, *25*, 100658.

[53] Melin, J.; Cano, S.J.; Gillman, A.; Marquis, S.; Flöel, A.; Göschel, L.; Pendrill, L.R. Traceability and comparability through crosswalks with the NeuroMET Memory Metric. *Scientific Reports*, **2023**, *13*, 5179.

[54] Pendrill, L.R. Man as a measurement instrument. *NCSLi Measure: The Journal of Measurement Science*, **2014**, *9*, 22-33.

[55] Pendrill, L.R. Using measurement uncertainty in decision-making and conformity assessment. *Metrologia*, **2014**, *51*, S206-S218.

[56] Pendrill, L.R. *Quality assured measurement: Unification across social and physical sciences*. Springer, Cham, Switzerland, 2019. Springer Series in Measurement Science and Technology.

[57] Pendrill, L.; Fisher, W.P. Jr. Counting and quantification: Comparing psychometric and metrological perspectives on visual perceptions of number. *Measurement*, **2015**, *71*, 46-55.

[58] Andersen, E.B. *Discrete statistical models with social science applications*. North-Holland, Amsterdam, 1980.

[59] Brogden, H.E. The Rasch model, the law of comparative judgment and additive conjoint measurement. *Psychometrika*, **1977**, *42*, 631-634.

[60] Fischer, G.H. On the existence and uniqueness of maximum-likelihood estimates in the Rasch model. *Psychometrika*, **1981**, *46*, 59-77.

[61] Fisher, W.P. Jr.; Wright, B.D., Eds. Applications of probabilistic conjoint measurement. *International Journal of Educational Research*, **1994**, *21*, 557-664.

[62] Narens, L.; Luce, R.D. Measurement: The theory of numerical assignments. *Psychological Bulletin*, **1986**, *99*, 166-180.

[63] Rasch, G. *Probabilistic models for some intelligence and attainment tests*. Danmarks Paedogogiske Institut, Copenhagen, Denmark, 1960.

[64] Rasch, G. On general laws and the meaning of measurement in psychology. In Neyman, J., ed. *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability: Volume IV: Contributions to biology and problems of medicine*. University of California Press, Berkeley, California, 1961, 321-333 [http://www.rasch.org/memo1960.pdf].

[65] Rasch, G. On specific objectivity: An attempt at formalizing the request for generality and validity of scientific statements. *Danish Yearbook of Philosophy*, **1977**, *14*, 58-94.

[66] Wright, B.D. Solving measurement problems with the Rasch model. *Journal of Educational Measurement*, **1977**, *14*, 97-116 [http://www.rasch.org/memo42.htm].

[67] Wright, B.D. Fundamental measurement for outcome evaluation. *Physical Medicine & Rehabilitation State of the Art Reviews*, **1997**, *11*, 261-288.

[68] Pendrill, L.R. Discussion on the definition of the term "unit". In Ulrich, J., ed. *pp. 12-16, in Report of the 24th meeting of the Consultative Committee for Units to the International Committee for Weights and Measures*. Pavillon de Breteuil, Sèvres, France. Bureau International des Poids et Mesures, Consultative Committee for Units (CCU), 2019.

[69] Hutchins, E. Cognitive ecology. *Topics in Cognitive Science*, **2010**, *2*, 705-715.

[70] Hutchins, E. The cultural ecosystem of human cognition. *Philosophical Psychology*, **2014**, *27*, 34-49.

[71] Latour, B. How to talk about the body? The normative dimension of science studies. *Body & Society*, **2004**, *10*, 205-229.

[72] Nersessian, N.J. Engineering concepts: The interplay between concept formation and modeling practices in bioengineering sciences. *Mind, Culture, and Activity*, **2012**, *19*, 222-239.

[73] Overton, W.F. Contexts of meaning: The computational and the embodied mind. In Overton, W.; D. Palermo, D., Eds. *The nature and ontogenesis of meaning*. Lawrence Erlbaum Associates, Inc., Hillsdale, New Jersey, USA, 1994, 1-18.

[74] Overton, W.F. Beyond dichotomy: An embodied active agent for cultural psychology. *Culture & Psychology*, **1997**, *3*, 315-334.

[75] Overton, W.F. Understanding, explanation, and reductionism: Finding a cure for Cartesian anxiety. In Smith, L.; Brown, T., Eds. *Reductionism and the development of knowledge*. Erlbaum, Mahwah, New Jersey, USA, 2002, 29-51.

[76] Overton, W.F. A coherent metatheory for dynamic systems: Relational organicism-contextualism. *Human Development*, **2007**, *50*, 154-159.

[77] Garcia, R. The structure of knowledge and the knowledge of structure. In Beilin, H.; Pufall, P., Eds. *Piaget's theory: Prospects and possibilities*. Lawrence Erlbaum Associates, Inc., Hillsdale, New Jersey, USA, 1992, 21-38.

[78] Overton, W.F. The arrow of time and the cycle of time: Concepts of change, cognition, and embodiment. *Psychological Inquiry*, **1994**, *5*, 215-237.

[79] Akrich, M.; Callon, M.; Latour, B. The key to success in innovation Part I: The art of interessement. *International Journal of Innovation Management*, **2002**, *6*, 187-206 [doi: 10.1142/S1363919602000550].

[80] Akrich, M.; Callon, M.; Latour, B. The key to success in innovation Part II: The art of choosing a good spokesperson. *International Journal of Innovation Management*, **2002**, *6*, 207-225.

[81] Blok, A.; Farias, I.; Roberts, C. Eds. *The Routledge companion to Actor-Network Theory*. Routledge, London, UK, 2020.

[82] Latour, B. *Reassembling the social: An introduction to Actor-Network-Theory*. Oxford University Press, Oxford, UK, 2005.

[83] Latour, B. To modernise or ecologise? That is the question. In Braun, B.; Castree, N., Eds., *Remaking reality: Nature at the millennium*. Routledge, London, UK, 1998, 221-242.

[84] Fisher, W.P. Jr.; Stenner, A.J. Ecologizing vs modernizing in measurement and metrology. *Journal of Physics Conference Series*, **2018**, *1044*,[http://iopscience.iop.org/article/10.1088/1742-6596/1044/1/012025].

[85] Morrison, J.; Fisher, W.P. Jr, Caliper: Measuring success in STEM learning ecosystems. *Measurement: Sensors*, **2021**, *18*, 100327.

[86] Piaget, J. *Play, dreams and imitation in childhood (Vol. 25)*. Routledge, New York, NY, USA 2013.

[87] Overton, W.F. Developmental psychology: Philosophy, concepts, and methodology. *Handbook of Child Psychology*, **1998**, *1*, 107-188.

[88] Overton, W.F., Processes, relations and Relational-Developmental-Systems. In Overton, W.F.; Molenaar, P.C.M., Eds. *Theory and Method. Volume 1 of the Handbook of child psychology and developmental science (7th Ed.)*. Wiley, Hoboken, New Jersey, USA, 2015, 9-62.

[89] Gadamer, H.-G. *Truth and method*. Crossroad, New York, NY, USA 1989.

[90] Commons, M.L.; Bresette, L.M. Illuminating major creative scientific innovators with postformal stages. In Hoare, C., ed. *Handbook of adult development and learning*. Oxford University Press, Oxford, UK, 2006, 255-280.

[91] Commons, M.L.; Ross, S.N.; Bresette, L.M. The connection between postformal thought, stage transition, persistence, and ambition and major scientific innovations. In Artistico, D.; Berry, J.; Black, J.; Cervone, D.; Lee, C.; Orom, H., Eds. *The Oxford handbook of reciprocal adult development and learning*. Oxford University Press, Oxford, UK, 2011, 287-301.

[92] Nersessian, N.J. Child's play. *Philosophy of Science*, **1996**, *63*, 542-546.

[93] Nuzzo, A. Thinking being: Method in Hegel's logic of being. In Houlgate, S.; Bauer, M., Eds. *A Companion to Hegel*. Blackwell, Oxford, UK, 2011, 109-138.

[94] Fischer, K.W. A theory of cognitive development: The control and construction of hierarchies of skills. *Psychological Review*, **1980**, *87*, 477- 531.

[95] Dawson, T.L.; Fischer, K.W.; Stein, Z. Reconsidering qualitative and quantitative research approaches: A cognitive developmental perspective. *New Ideas in Psychology*, **2006**, *24*, 229-239.

[96] Fischer, K.W.; Farrar, M.J. Generalizations about generalization: How a theory of skill development explains both generality and specificity. *International Journal of Psychology*, **1987**, *22*, 643-677.

[97] Barab, S.A.; Plucker, J.A. Smart people or smart contexts? Cognition, ability, and talent development in an age of situated approaches to knowing and learning. *Educational Psychologist*, **2002**, *37*, 165-182.

[98] Shore, D.A. *Launching and leading change initiatives in health care organizations: Managing successful projects (Vol. 213)*. Jossey-Bass, San Francisco, California, USA2014.

[99] Commons, M.L.; Goodheart, E.A. Cultural progress is the result of developmental level of support. *World Futures: The Journal of New Paradigm Research*, **2008**, *64*, 406-415.

[100] Dewey, J. *The public and its problems*. Swallow Press, Ohio University Press, Athens, Ohio, 1954.

[101] Whitehead, A.N. *Science and the modern world*. Macmillan, New York, NY, USA 1925.

[102] Bateson, G. *Steps to an ecology of mind: Collected essays in anthropology, psychiatry, evolution, and epistemology*. University of Chicago Press, Chicago, Illinois, USA, 1972.

[103] Fisher, R.A. Indeterminism and natural selection. *Philosophy of Science*, **1934**, *1*, 99-117.

[104] Odum, H.T.; Cantlon, J.E.; Kornicker, L.S. An organizational hierarchy postulate for the interpretation of species-individual distributions, species entropy, ecosystem evolution, and the meaning of a species-variety index. *Ecology*, **1960**, *41*, 395-399.

[105] Acs, Z.J.; Estrin, S.; Mickiewicz, T.; Szerb, L. Entrepreneurship, institutional economics, and economic growth: An ecosystem perspective. *Small Business Economics*, **2018**, *51*, 501-514.

[106] Latour, B. *Science in action: How to follow scientists and engineers through society*. Harvard University Press, New York, NY, USA 1987.

[107] Mitchell, T. Rethinking economy. *Geoforum*, **2008**, *39*, 1116-1121.

[108] Fisher, W.P. Jr. What the world needs now: A bold plan for new standards. *Standards Engineering*, **2012**, *64*, 1 & 3-5 [http://ssrn.com/abstract=2083975].

[109] Fisher, W.P. Jr. Measure and manage: Intangible assets metric standards for sustainability. In Marques, J.; Dhiman, S.; Holt, S., Eds. *Business administration education: Changes in management and leadership strategies*. Palgrave Macmillan, New York, NY, USA 2012, 43-63.

[110] Fisher, W.P. Jr. Measurement systems, brilliant results, and brilliant processes in healthcare: Untapped potentials of person-centered outcome metrology for cultivating trust. In Fisher, W.P. Jr.; Cano, S., Eds. *Person-centered outcome metrology: Principles and applications for high stakes decision making*. Springer, Cham, Switzerland, 2023, 357-396.

[111] Fisher, W.P. Jr. Separation theorems in econometrics and psychometrics: Rasch, Frisch, two Fishers, and implications for measurement. *Journal of Interdisciplinary Economics*, **2023**, *35*, 29-60.

[112] Lenton, T.M.; Latour, B. Gaia 2.0: Could humans add some level of self-awareness to Earth's self-regulation? *Science*, **2018**, *361*, 1066-1068.

[113] Lerner, R.M.; Overton, W.F. Reduction to absurdity: Why epigenetics invalidates all models involving genetic reduction. *Human Development*, **2017**, *60*, 107-123.

[114] Meloni, M. *Impressionable biologies: From the archaeology of plasticity to the sociology of epigenetics*. Routledge, New York, NY, USA 2019.

[115] Tadelis, S.; Williamson, O. Transaction cost economics. Haas School of Business, University of California, Berkeley, 2010.

[116] Hayek, F.A. The use of knowledge in society. *American Economic Review*, **1945**, *35*, 519-530. (Rpt. in F. A. Hayek, *Individualism and economic order*. University of Chicago Press, Chicago, Illinois, USA. 1948, 77-91.)

[117] Fisher, W.P. Jr., The central theoretical problem of the social sciences. *Rasch Measurement Transactions*, **2014**, *28*,1464-1466.

[118] Prigogine, I. Science, civilization and democracy: Values, systems, structures and affinities. *Futures*, **1986**, *18*, 493-507.

[119] Prigogine, I. *The end of certainty: Time, chaos, and the new laws of nature*. Free Press, New York, NY, USA 1997.

[120] Galison, P. Trading zone: Coordinating action and belief. In Biagioli, M., ed. *The science studies reader*. Routledge, New York, NY, USA 1999, 137-160.

[121] Galison, P. *Image and logic: A material culture of microphysics*. University of Chicago Press, Chicago, Illinois, USA, 1997.

[122] Brier, S. Cybersemiotics: A new foundation for transdisciplinary theory of information, cognition, meaningful communication and the interaction between nature and culture. *Integral Review*, **2013**, *9*, 220-263.

[123] Brier, S. Cybersemiotics and the reasoning powers of the universe: Philosophy of information in a semiotic-systemic transdisciplinary approach. *Green Letters: Studies in Ecocriticism*, **2015**, *19*, 280-292.

[124] Brier, S. Cybersemiotic systemic and semiotical based transdisciplinarity. In Vidales, C.; Brier, S., Eds. *Introduction to cybersemiotics: A transdisciplinary perspective*. Springer, Cham, Switzerland, 2021, 17-31.

[125] Danesi, M. Semiotics as a metalanguage for the sciences. In Bankov, K.; Cobley, P., Eds. *Semiotics and its masters*. DeGruyter, Berlin, 2017, 61-81.

[126] Nöth, W. The semiotics of models. *Sign Systems Studies*, **2018**, *46*, 7-43.

[127] Nöth, W. System, sign, information, and communication in cybersemiotics, systems theory, and Peirce. In Vidales, C.; Brier, S., Eds. *Introduction to cybersemiotics: A transdisciplinary perspective*. Springer, Cham, Switzerland, 2021, 75-95.

[128] Olteanu, A.; Campbell, C.; Feil, S. Naturalizing models: New perspectives in a Peircean key. *Biosemiotics*, **2020**, *13*, 179-197.

[129] Sebeok, T.A. *Signs: An introduction to semiotics*. University of Toronto Press, Toronto, 2001.

[130] Ackermann, J.R. *Data, instruments, and theory: A dialectical approach to understanding science*. Princeton University Press, Princeton, New Jersey, 1985.

[131] Blok, A.; Nakazora, M.; Winthereik, B.R. Infrastructuring environments. *Science as Culture*, **2016**, *25*, 1-22.

[132] Bowker, G.C. How knowledge infrastructures learn. In Harvey, P.; Jensen, C.B.; Morita, A., Eds. *Infrastructures and social complexity: A companion*. Routledge, New York, NY, USA 2016, 391-403.

[133] Bowker, G.C.; Star, S.L. How things (actor-net) work: Classification, magic and the ubiquity of standards. *Philosophia*, **1996**, *25*, 195-220.

[134] Hankins, T.L.; Silverman, R.J. *Instruments and the imagination*. Princeton University Press, Princeton, New Jersey, 1999.

[135] Heelan, P.A., Natural science as a hermeneutic of instrumentation. *Philosophy of Science*, **1983**, *50*, 181-204.

[136] Heelan, P.A. Paradoxes of measurement. *Annals of the New York Academy of Sciences*, **2003**, *988*, 114-127.

[137] Ihde, D. *Instrumental realism: The interface between philosophy of science and philosophy of technology*. Indiana University Press, Bloomington, Indiana, 1991. The Indiana Series in the Philosophy of Technology.

[138] Knorr Cetina, K. Laboratory studies: The cultural approach to the study of science. In Jasanoff, S.; Markle, G.E.; Petersen, J.C.; Pinch, T., Eds. *Handbook of science and technology studies*. Sage Publications, Thousand Oaks, California, 1995, 140-166.

[139] Knorr Cetina, K. *Epistemic cultures: How the sciences make knowledge*. Harvard University Press, Cambridge, Massachusetts, 1999.

[140] Latour, B. Give me a laboratory and I will raise the world. In Knorr-Cetina, K.D.; Mulkay, M., Eds. *Science observed: Perspectives on the social study of science*. Sage Publications, London, England, 1983, 141-170.

[141] Latour, B. Postmodern? No, simply amodern: Steps towards an anthropology of science. *Studies in History and Philosophy of Science*, **1990**, *21*, 145-171.

[142] Latour, B. *We have never been modern*. Harvard University Press, Cambridge, Massachusetts, 1993.

[143] Latour, B. Love your monsters: Why we must care for our technologies as we do our children, *Breakthrough Journal*, **2012**, *2*, Available: https://thebreakthrough.org/journal/issue-2/love-your-monsters.

[144] Rabkin, Y.M. Rediscovering the instrument: Research, industry, and education. In Bud, R.; Cozzens, S.E., Eds. *Invisible connections: Instruments, institutions, and science*. SPIE Optical Engineering Press, Bellingham, Washington, 1992, 57-82.

[145] Rodríguez-Giralt, I.; Marrero-Guillamón, I.; Milstein, D. Reassembling activism, activating assemblages: An introduction. *Social Movement Studies*, **2018**, *17*, 257-268.

[146] Schaffer, S. Late Victorian metrology and its instrumentation: A manufactory of Ohms. In Bud, R.; Cozzens, S.E., Eds. *Invisible connections: Instruments, institutions, and science*. SPIE Optical Engineering Press, Bellingham, WA, 1992, 23-56.

[147] Schaffer, S. Metrology, metrication, and Victorian values. In Lightman, B. ed. *Victorian science in context*. University of Chicago Press, Chicago, Illinois, USA, 1997, 438-474.

[148] Schaffer, S. Modernity and metrology. In Guzzetti, L., Ed. *Science and power: The historical foundations of research policies in Europe*. Office for Official Publications of the European Communities, Luxembourg, 2000, 71-91.

[149] Shapin, S. The invisible technician. *American Scientist*, **1989**, *77*, 554-563.

[150] Shapin, S.; Schaffer, S. *Leviathan and the air-pump: Hobbes, Boyle, and the experimental life*. Princeton University Press, Princeton, New Jersey, USA, 1985.

[151] Vertesi, J.; Ribes, D. *DigitalSTS: A field guide for science & technology studies*. Princeton University Press, Princeton, New Jersey, USA, 2019.

[152] Vonderau, A. Scaling the cloud: Making state and infrastructure in Sweden. *Ethnos Journal of Anthropology*, **2019**, *84*, 698-718.

[153] Guttman, L. The language of science. In Levy, S., ed. *Louis Guttman on theory and methodology: Selected writings*. Dartmouth Publishing Company, Brookfield, VT, 1994, 79-102.

[154] Wright, B.D. A history of social science measurement. *Educational Measurement: Issues and Practice*, **1997**, *16*, 33-45, 52 [http://www.rasch.org/memo62.htm].

[155] Stenner, A.J.; Horabin, I. Three stages of construct definition. *Rasch Measurement Transactions*, **1992**, *6*, 229 [http://www.rasch.org/rmt/rmt63b.htm].

[156] Dawson, T.L. Rethinking educational assessment in light of a strong theory of development. In Mascolo, M.F.; Bidell, T.R., Eds. *Handbook of integrative developmental science: Essays in honor of Kurt W. Fischer*. Routledge, New York, NY, USA 2020, 423-450.

[157] Chaitin, G.J. Randomness and complexity in pure mathematics. *International Journal of Bifurcation and Chaos*, **1994**, *4*, 3-15.

[158] Calude, C.S. Incompleteness, complexity, randomness and beyond. *Minds and Machines*, **2002**, *12*, 503-517.

[159] Calude, C.S. Ed. *Randomness & complexity from Leibniz to Chaitin*. World Scientific, Singapore, 2007.

[160] Chaitin, G.J. Randomness in arithmetic and the decline & fall of reductionism in pure mathematics. *EATCS Bulletin*, **1993**, *50*, 314-328.

[161] Chaitin, G.J. *The limits of mathematics*. Springer-Verlag, London, 2003.

[162] Dewey, J. *The quest for certainty: A study of the relation of knowledge and action*. Perigee Books, G. P. Putnam's Sons, New York, NY, USA 1929.
[163] Kline, M. *Mathematics: The loss of certainty*. Oxford University Press, Oxford, UK, 1980.
[164] Nagel, E.; Newman, J.R. *Gödel's Proof.* New York University Press, New York, NY, USA 1958.
[165] Weinsheimer, J. *Gadamer's hermeneutics: A reading of Truth and Method*. Yale University Press, New Haven, Connecticut, USA, 1985.
[166] Kuhn, T.S. *The structure of scientific revolutions*. University of Chicago Press, Chicago, Illinois, USA, 1970.
[167] Dupre, J. *The disorder of things: Metaphysical foundations of the disunity of science*. Harvard University Press, Cambridge, Massachusetts, 1993.
[168] Galison, P.; Stump, D. eds. *Disunity and contextualism*. Stanford University Press, Palo Alto, California, USA, 1996.
[169] Star, S.L.; Griesemer, J.R. Institutional ecology, 'translations,' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science*, **1989**, *19*, 387-420.
[170] Woolley, A.W.; Fuchs, E. Collective intelligence in the organization of science. *Organization Science*, **2011**, *22*, 1359-1367.
[171] Cripps, S.; Durrant-Whyte, H. Uncertainty: Nothing is more certain. *Environmetrics*, **2023**, *34*,e2745.
[172] Commons, M.L.; Richards, F.A. Four postformal stages. In Demick, J.; Andreoletti, C., Eds. *Handbook of adult development*. Plenum Press, New York, NY, USA 2002, 199-219.
[173] Commons, M.L.; Richards, F.A.; Kuhn, D. Systematic and metasystematic reasoning: A case for levels of reasoning beyond Piaget's stage of formal operations. *Child Development*, **1982**, *53*, 1058-1069.
[174] Commons, M.L.; Trudeau, E.J.; Stein, S.A.; Richards, F.A.; Krause, S.R. Hierarchical complexity of tasks shows the existence of developmental stages. *Developmental Review*, **1998**, *18*, 237-278.
[175] Berti, M.; Simpson, A.V. The dark side of organizational paradoxes: The dynamics of disempowerment. *Academy of Management Review*, **2021**, *46*, 252-274.
[176] Ross, S.N. A future society functioning at the paradigmatic stage? *World Futures*, **2008**, *64*, 554-562.
[177] Bernstein, W.J. *The birth of plenty: How the prosperity of the modern world was created*. McGraw-Hill, New York, NY, USA 2004.
[178] De Soto, H. *The mystery of capital: Why capitalism triumphs in the West and fails everywhere else*. Basic Books, New York, NY, USA 2000.
[179] Ekins, P. A four-capital model of wealth creation. In Ekins, P.; Max-Neef, M., Eds. *Real-life economics: Understanding wealth creation*. Routledge, London, 1992, 147-155.
[180] Ekins, P.; Dresner, S.; Dahlstrom, K. The four-capital method of sustainable development evaluation. *European Environment*, **2008**, *18*, 63-80.
[181] Fisher, W.P. Jr. Invariance and traceability for measures of human, social, and natural capital: Theory and application. *Measurement*, **2009**, *42*, 1278-1287.
[182] Prigogine, I. Dialogue avec Piaget sur l'irréversible [Dialogue with Piaget concerning the irreversible]. *Archives de Psychologie*, **1982**, *50*, 7-16.
[183] Prigogine, I. Laws of nature and human conduct: Specificities and unifying themes. *Mondes en développement*, **1986**, *14*, 5-8.
[184] Brent, S.B. Prigogine's model for self-organization in nonequilibrium systems: Its relevance for developmental psychology. *Human Development*, **1978**, *21*, 374-387.
[185] Brent, S.B. *Psychological and social structures*. Routledge.
[186] Chapman, M. *Constructive evolution: Origins and development of Piaget's thought*. Cambridge University Press.
[187] Ståhle, P.; Mononen, L.; Tynjälä, P.; Kallio, E.K. Systems thinking and adult cognitive development. In Kallio, E.K., ed. *Development of adult thinking: Interdisciplinary perspectives on cognitive development and adult learning*. Routledge, London, 2020, 191-207.
[188] Prigogine, I.; Allen, P.M. The challenge of complexity. In Schieve, W.C.; Allen, P., Eds. *Self-organization and dissipative structures: Applications in the physical and social sciences*. University of Texas Press, Austin, Texas, 1982, 3-39.
[189] Fisher, R.A. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London, A*, **1922**, *222*, 309-368.
[190] Fisher, R.A. Two new properties of mathematical likelihood. *Proceedings of the Royal Society, A*, **1934**, *144*, 285-307.
[191] Duncan, O.D.; Stenbeck, M. Panels and cohorts: Design and model in the study of voting turnout. In Clogg, C.C., ed. *Sociological Methodology 1988*. American Sociological Association, Washington, DC, 1988, 1-35.
[192] Duncan, O.D. What if? *Contemporary Sociology*, **1992**, *21*, 667-668.
[193] Andrich, D. Distinctions between assumptions and requirements in measurement in the social sciences. In Keats, J.A.; Taft, R.; Heath, R.A.; Lovibond, S.H., Eds. *Mathematical and Theoretical Systems:*

Proceedings of the 24th International Congress of Psychology of the International Union of Psychological Science, Vol. 4. Elsevier Science Publishers, North-Holland, 1989, 7-16.

[194] Cohen, J. The earth is round ($p < 0.05$). *American Psychologist*, **1994**, *49*, 997-1003.

[195] Guttman, L. The illogic of statistical inference for cumulative science. *Applied Stochastic Models and Data Analysis*, **1985**, *1*, 3-10.

[196] Meehl, P.E. Theory-testing in psychology and physics: A methodological paradox. *Philosophy of Science*, **1967**, *34*, 103-115.

[197] Michell, J. Measurement scales and statistics: A clash of paradigms. *Psychological Bulletin*, **1986**, *100*, 398-407.

[198] Molenaar, P. A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research and Perspective*, **2004**, *2*, 201-218.

[199] Rodgers, J.L. The epistemology of mathematical and statistical modeling: A quiet methodological revolution. *American Psychologist*, **2010**, *65*, 1-12.

[200] Rogosa, D. Casual [sic] models do not support scientific conclusions: A comment in support of Freedman. *Journal of Educational Statistics*, **1987**, *12*, 185-195.

[201] Wilson, M.R. Seeking a balance between the statistical and scientific elements in psychometrics. *Psychometrika*, **2013**, *78*, 211-236.

[202] Wright, B.D. Despair and hope for educational measurement. *Contemporary Education Review*, **1984**, 3, 281-288 [http://www.rasch.org/memo41.htm].

[203] Wright, B.D. Fundamental measurement for psychology. In Embretson, S.E.; Hershberger, S.L., Eds. *The new rules of measurement: What every educator and psychologist should know*. Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1999, 65-104 [http://www.rasch.org/memo64.htm].

[204] Guttman, L. Measurement as structural theory. *Psychometrika*, **1971**, *36*, 329-347.

[205] Bich, W. Measurement uncertainty and metrological traceability. In Jeckelmann, B.; Edelmaier, R., Eds. *Metrological infrastructure*. De Gruyter Oldenbourg, Berlin, Germany, 2023, 23-56.

[206] Joint Committee for Guides in Metrology, W.G.1. *Guide to the expression of uncertainty in measurement -- Part 6: Developing and using measurement models*. International Bureau of Weights and Measures--BIPM, Sevres, France, 2020.

[207] Rossi, G. Probability in metrology. In Pavese, F.; Forbes, A.B., Eds. *Data modeling for metrology and testing in measurement science*. Birkhäuser, Boston, Massachusetts, USA, 2009, 31-70.

[208] Rossi, G.B.; Crenna, F. A formal theory of the measurement system. *Measurement*, **2018**, *116*, 644-651.

[209] Morawski, R.Z. An application-oriented mathematical meta-model of measurement. *Measurement*, **2013**, *46*, 3753-3765.

[210] Rasch, G. An individualistic approach to item analysis. In Lazarsfeld, P.F.; Henry, N.W., Eds. *Readings in mathematical social science*. Science Research Associates, Chicago, Illinois, USA, 1966, 89-108.

[211] Luecht, R.M.; Hambleton, R.K. Item response theory: A historical perspective and brief introduction to applications. In Clauser, B.E.; Bunch, M.B., Eds. *The history of educational measurement: Key advancements in theory, policy, and practice*. Routledge.

[212] Andrich, D. Constructing fundamental measurements in social psychology. In Keats, J.A.; Taft, R.; Heath, R.A.; Lovibond, S.H., Eds. *Mathematical and theoretical systems. Proceedings of the 24th International Congress of Psychology of the International Union of Psychological Science, Vol. 4*. North-Holland, Amsterdam, Netherlands, 1989, 17-26.

[213] Embretson, S.E. Item Response Theory models and spurious interaction effects in factorial ANOVA designs. *Applied Psychological Measurement*, **1996**, *20*, 201-212.

[214] Hambleton, R.; Wright, B.D.; Crocker, L.; Masters, G.; van der Linden, W. Hambleton's 9 theses. *Rasch Measurement Transactions*, **1992**, *6*, 215 [http://www.rasch.org/rmt/rmt62.htm].

[215] Lumsden, J. Tests are perfectly reliable. *British Journal of Mathematical and Statistical Psychology*, **1978**, *31*, 19-26.

[216] San Martin, E.; Rolin, J.M. Identification of parametric Rasch-type models. *Journal of Statistical Planning and Inference*, **2013**, *143*, 116-130.

[217] San Martin, E.; Gonzalez, J.; Tuerlinckx, F. On the unidentifiability of the fixed-effects 3 PL model. *Psychometrika*, **2015**, *80*, 450-467.

[218] Verhelst, N.D.; Glas, C.A.W. The one parameter logistic model. In Fischer, G.H.; Molenaar, I.W., Eds. *Rasch models: Foundations recent developments, and applications*. Springer, New York, NY, USA 1995, 215-237.

[219] Linacre, J.M. What is Item Response Theory, IRT? A tentative taxonomy. *Rasch Measurement Transactions*, **2003**, *17*, 926-927.

[220] Hambleton, R.K.; Swaminathan, H.; Rogers, L. *Fundamentals of item response theory*. Sage Publications, Newbury Park, California, 1991.

[221] Shrader-Frechette, K. *Tainted: How philosophy of science can expose bad science*. Oxford University Press, New York, NY, USA 2014.

[222] Kuhn, T.S. *The essential tension: Selected studies in scientific tradition and change*. University of Chicago Press, Chicago, Illinois, USA, Illinois, 1977.

[223] Butterfield, H. *The origins of modern science (revised edition)*. The Free Press, New York, NY, USA 1957.

[224] Kant, I. *Critique of pure reason*, Smith, N.K., trans. St. Martin's Press, New York, NY, USA 1929/1965.

[225] Boumans, M. Paul Ehrenfest and Jan Tinbergen: A case of limited physics transfer. In De Marchi, N., ed. *Non-natural social science: Reflecting on the enterprise of "More Heat than Light"*. Duke University Press, Durham, North Carolina, USA, 1993, 131-156.

[226] Boumans, M. Measurement in economic systems. *Measurement*, **2005**, *38*, 275-284.

[227] Brenner-Golomb, N. R. A. Fisher's philosophical approach to inductive inference. In Keren, G.; Lewis, C., Eds. *A handbook for data analysis in the behavioral sciences: Methodological issues*. Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1993, 283-307.

[228] Wright, B.D. Useful measurement through concurrent equating and one-step (concurrent) item banking. *Rasch Measurement Transactions*, **1988**, *2*, 24 [http://www.rasch.org/rmt/rmt22f.htm].

[229] Fisher, W.P. Jr. Contextualizing sustainable development metric standards: Imagining new entrepreneurial possibilities. *Sustainability*, **2020**, *12*, 1-22.

[230] Wilson, M.; Fisher, W.P. Jr. Eds. *Psychological and social measurement: The career and contributions of Benjamin D. Wright*, Cain, M.G.; Rossi, G.B.; Tesai, J.; van Veghel, M.; Jhang, K.-Y., Eds. Springer Nature, Cham, Switzerland, Switzerland, 2017. Springer Series in Measurement Science and Technology.

[231] Perline, R.; Wright, B.D.; Wainer, H. The Rasch model as additive conjoint measurement. *Applied Psychological Measurement*, **1979**, *3*, 237-255.

[232] Newby, V.A.; Conner, G.R.; Grant, C.P.; Bunderson, C.V. The Rasch model and additive conjoint measurement. *Journal of Applied Measurement*, **2009**, *10*, 348-354.

[233] Peabody, M.R.; Muckle, T.J.; Meng, Y. Applying a mixture Rasch model-based approach to standard setting. *Educational Measurement: Issues and Practice*, **2023**, *42*, 5-12.

[234] Smith, R.M.; Julian, E.; Lunz, M.; Stahl, J.; Schulz, M.; Wright, B.D. Applications of conjoint measurement in admission and professional certification programs. *International Journal of Educational Research*, **1994**, *21*, 653-664.

[235] Black, P.; Wilson, M.; Yao, S. Road maps for learning: A guide to the navigation of learning progressions. *Measurement: Interdisciplinary Research and Perspectives*, **2011**, *9*, 1-52.

[236] Duckor, B.; Holmberg, C. Exploring how to model formative assessment trajectories of posing-pausing-probing practices: Toward a teacher learning progressions framework for the study of novice teachers. *Journal of Educational Measurement*, **2019**, *56*, 836-890.

[237] Duckor, B.; Holmberg, C. Examining conceptual frameworks for understanding teacher learning in the domain of formative assessment. In *International Encyclopedia of Education*, Volume 5. Elsevier, 2023, 581-595.

[238] Duckor, B.; Holmberg, C. *Feedback for continuous improvement in the classroom: New perspectives, practices, and possibilities*. Corwin, 2023.

[239] Wilson, M.R. Cognitive diagnosis using item response models. *Zeitschrift für Psychologie/Journal of Psychology (Special Issue: Current Issues in Competence Modeling and Assessment)*, **2008**, *216*, 74-88.

[240] Wilson, M.R. Measuring progressions: Assessment structures underlying a learning progression. *Journal of Research in Science Teaching*, **2009**, *46*, 716-730.

[241] Wilson, M.R. Making measurement important for education: The crucial role of classroom assessment. *Educational Measurement: Issues and Practice*, **2018**, *37*, 5-20.

[242] Commons, M.L. What are the relationships between four notions of stage change? *Behavioral Development Bulletin*, **2014**, *19*, 15-21.

[243] Commons, M.L.; Goodheart, E.A.; Pekker, A.; Dawson-Tunik, T.L.; Adams, K.M. Using Rasch scaled stage scores to validate orders of hierarchical complexity of balance beam task sequences. *Journal of Applied Measurement*, **2008**, *9*, 182-199.

[244] Commons, M.L.; Miller, P.M.; Ramakrishnan, S.; Giri, S. Employee management using behavioral development theory. *Behavioral Development*, **2018**, *23*, 22-33.

[245] Dawson, T.L.; Goodheart, E.A.; Draney, K.; Wilson, M.; Commons, M.L. Concrete, abstract, formal, and systematic operations as observed in a "Piagetian" balance-beam task. In Garner, M.; Engelhard, G. Jr.; Fisher, W.P. Jr.; Wilson, M., Eds. *Advances in Rasch Measurement, Vol. 1*. JAM Press, Maple Grove, MN, USA 2010, 572-590.

[246] Dawson-Tunik, T.L.; Commons, M.; Wilson, M.; Fischer, K. The shape of development. *The European Journal of Developmental Psychology*, **2005**, *2*, 163-196.

[247] Draney, K.; Wilson, M. Application of the polytomous Saltus model to stage-like proportional reasoning data. In Van der Ark, L.A.; Croon, M.A.; Sijtsma, K., Eds. *New developments in categorical data analysis for the social and behavioral sciences*. Psychology Press, 207-226.

[248] Draney, K.; Wilson, M. Application of the Saltus model to stage-like data: Some applications and current developments. In von Davier, M., ed. *Multivariate and mixture distribution Rasch models*. Springer, New York, NY, USA 2006.

[249] Mueller, U.; Overton, W.F. Thinking about thinking - thinking about measurement: A Rasch analysis of recursive thinking. *Journal of Applied Measurement*, **2010**, *11*, 78-90.

[250] Wilson, M.R. Saltus: A psychometric model of discontinuity in cognitive development. *Psychological Bulletin*, **1989**, *105*, 276-289.

[251] Barney, M. Transdisciplinary measurement through AI: Hybrid metrology and psychometrics powered by Large Language Models. In Fisher, W.P. Jr.; Pendrill, L. eds. *Models, measurement, and metrology extending the SI*. De Gruyter, Berlin, Germany, 2024.

[252] Barney, M.; Fisher, W.P. Jr. Adaptive measurement and assessment. *Annual Review of Organizational Psychology and Organizational Behavior*, **2016**, *3*, 469-490.

[253] Lunz, M.E.; Bergstrom, B.A.; Gershon, R.C. Computer adaptive testing. *International Journal of Educational Research*, **1994**, *21*, 623-634.

[254] Cano, S.; Pendrill, L.; Melin, J.; Fisher, W.P. Jr. Towards consensus measurement standards for patient-centered outcomes. *Measurement*, **2019**, *141*, 62-69.

[255] Fisher, W.P. Jr.; Stenner, A.J. Theory-based metrological traceability in education: A reading measurement network. *Measurement*, **2016**, *92*, 489-496.

[256] Masters, G.N. Special issue: Programme for International Student Assessment (PISA). *Journal of Applied Measurement*, **2007**, *8*, 235-335.

[257] Quaglia, M.; Pendrill, L.; Melin, J.; Cano, S.; 18HLT09 NeuroMET2 Consortium *Publishable Summary for 18HLT09 NeuroMET2: Metrology and innovation for early diagnosis and accurate stratification of patients with neurodegenerative diseases*. EMPIR NeuroMET. EURAMET, Teddington, Middlesex, UK, 2019-2022, 5 pp.

[258] Stenner, A.J. The Lexile Framework: A common metric for matching readers and texts. *California School Library Journal*, **2001**, *25*, 41-42.

[259] Aldrich, J. Autonomy. *Oxford Economic Papers*, **1989**, *41*, 15-34.

[260] Koopmans, T.C.; Reiersøl, O. The identification of structural characteristics. *The Annals of Mathematical Statistics*, **1950**, *XXI*, 165-181.

[261] Andrich, D.; Douglas, G.A. A binomial latent trait model for the study of Likert-style attitude questionnaires. *British Journal of Mathematical and Statistical Psychology*, **1978**, *31*, 84-98.

[262] Andrich, D. A rating formulation for ordered response categories. *Psychometrika*, **1978**, *43*, 561-573.

[263] Andrich, D. Sufficiency and conditional estimation of person parameters in the polytomous Rasch model. *Psychometrika*, **2010**, *75*, 292-308.

[264] Wright, B.D.; Masters, G.N. *Rating scale analysis*. MESA Press, Chicago, Illinois, USA, 1982.

[265] Graßhoff, U.; Holling, H.; Schwabe, R. D-optimal design for the Rasch counts model with multiple binary predictors. *British Journal of Mathematical and Statistical Psychology*, **2020**, *73*, 541-555.

[266] Meredith, W.M. *The Poisson distribution and Poisson process in psychometric theory*. Educational Testing Service, Princeton, New Jersey, USA, 1968, 81 pp.

[267] Rasch, G. The Poisson process as a model for a diversity of behavioral phenomena. Washington, DC. International Congress of Psychology.

[268] Andersen, E.B. Asymptotic properties of conditional maximum likelihood estimators. *Journal of the Royal Statistical Society*, **1970**, *30*, 283-301.

[269] Andersen, E.B. Sufficient statistics and latent trait models. *Psychometrika*, **1977**, *42*, 69-81.

[270] Barndorff-Nielsen, O. *Exponential families and conditioning*. Aarhus Universite, Denmark, 1973.

[271] Barndorff-Nielsen, O. *Information and exponential families in statistical theory*. John Wiley and Sons, New York, NY, USA 1978.

[272] Bradley, E.L. The equivalence of maximum likelihood and weighted least squares estimates in the exponential family. *Journal of the American Statistical Association*, **1973**, *68*, 199-200.

[273] Charnes, A.; Frome, E.L.; Yu, P.L. The equivalence of generalized least squares and maximum likelihood estimates in the exponential family. *Journal of the American Statistical Association*, **1976**, *71*, 169-171.

[274] Daum, F. The Fisher-Darmois-Koopman-Pitman theorem for random processes. Athens, Greece. 1986 25th IEEE Conference on Decision and Control.

[275] Fraser, D.A.S. On sufficiency and the exponential family. *Journal of the Royal Statistical Society: Series B (Methodological)*, **1963**, *25*, 115-123.

[276] Habermann, S. Maximum likelihood estimates in exponential response models. *The Annals of Statistics*, **1977**, *5*, 815-841.

[277] Koopman, B.O. On distributions admitting a sufficient statistic. *Transactions of the American Mathematical Society*, **1936**, *39*, 399-409.

[278] Linacre, J.M. *A user's guide to WINSTEPS Rasch-Model computer program, v. 5.6.2*. Winsteps.com, Beaverton, Oregon, USA, 2023.

[279] Gottlieb, G. *Individual development and evolution: The genesis of novel behavior*. Oxford University Press, Oxford, UK, 1992.

[280] Cole, M. Cultural mechanisms of cognitive development. In Amsel, E.; Renninger, K.A., Eds. *Change and development: Issues of theory, method, and application*. Lawrence Erlbaum, Mahwah, New Jersey, USA, 1997, 245-263.

[281] Overton, W.F. A new paradigm for developmental science: Relationism and relational-developmental systems. *Applied Developmental Science*, **2013**, *17*, 94-107.

[282] Wright, B.D.; Stone, M.H. *Best Test Design*. MESA Press, Chicago, Illinois, USA, 1979.

[283] Cano, S.; Pendrill, L.; Barbic, S.; Fisher, W.P. Jr. Patient-centred outcome metrology for healthcare decision-making. *Journal of Physics: Conference Series*, **2018**, *1044*,012057.

[284] Quaglia, M.; Pendrill, L.; Melin, J.; Cano, S.; 15HLT04 NeuroMET Consortium *Innovative measurements for improved diagnosis and management of neurodegenerative diseases*. EMPIR NeuroMET. EURAMET, Teddington, Middlesex, UK, 2016-2019, 36 pp.

[285] Petrosky, T.; Prigogine, I. Laws and events: The dynamical basis of self-organization. *Canadian Journal of Physics*, **1990**, *68*, 670-682.

[286] Bohm, D. A suggested interpretation of the quantum theory in terms of "hidden" variables. I. *Physical Review*, **1952**, *85*, 166-179.

[287] Bohm, D. A suggested interpretation of the quantum theory in terms of "hidden" variables. II. *Physical Review*, **1952**, *85*, 180-193.

[288] Bohm, D. *Wholeness and the implicate order*. Routledge & Kegan Paul, London, England, 1980.

[289] Bohm, D.; Hiley, B.J. Measurement understood through the quantum potential approach. *Foundations of Physics*, **1984**, *14*, 255-274.

[290] Bohm, D.; Hiley, B.J.; Kaloyerou, P.N. An ontological basis for the quantum theory. *Physics Reports*, **1987**, *144*, 321-375.

[291] Cowan, C.W.; Tumulka, R. Epistemology of wave function collapse in quantum physics. *The British Journal for the Philosophy of Science*, **2016**, *67*, 405-434.

[292] de Broglie, L.; Rosenfeld, L. Physics and metaphysics. *Nature*, **1958**, *181*, 1814-1814.

[293] Ellinor, L. Bohm's journey to dialogue: A look at its roots. In Banathy, B. Jenlink, ed. *Dialogue as a means of collective communication*. Springer, Boston, Massachusetts, USA, 2005, 255-276.

[294] Esfeld, M.; Lazarovici, D.; Hubert, M.; Dürr, D. The ontology of Bohmian mechanics. *The British Journal for the Philosophy of Science*, **2014**, *65*, 773-796.

[295] Goldstein, S. Quantum theory without observers-Part one. *Physics Today*, **1998**, *51*, 42-47.

[296] Goldstein, S. Quantum theory without observers-Part two. *Physics Today*, **1998**, *51*, 38-42.

[297] le Breton, P. Scholarly heroism: A transdisciplinary perspective on David Bohm's transformative scholarship and intellectual exile. *Journal of Humanistic Psychology*, **2018**, *58*, 415-430.

[298] Lovejoy, D. Objectivity, causality, and ideology in modern physics. *Science & Society*, **1999**, *63*, 433-468.

[299] Matarese, V. *The metaphysics of Bohmian mechanics: A comprehensive guide to the different interpretations of Bohmian ontology*, Esfeld, M.; Hartmann, S.; Newen, A., Eds. De Gruyter. Epistemic studies, 51.

[300] Walleczek, J.; Grössing, G.; Pylkkänen, P.; Hiley, B. Emergent quantum mechanics: David Bohm centennial perspectives. *Entropy*, **2019**, *21*, 113.

[301] Wheeler, J.A. The universe as a home for man. *American Scientist*, **1974**, *62*, 683-691.

[302] Wheeler, J.A. The participatory universe. *Science*, **1981**, *81*, 5.

[303] Wheeler, J.A. World as system self-synthesized by quantum networking. *IBM Journal of Research and Development*, **1988**, *32*, 4-15.

[304] Wheeler, J.A. Information, physics, quantum: The search for links. In *Feynman and computation: Exploring the limits of computers*. CRC Press, 309-336.

[305] Meijer, D.K. The universe as a cyclic organized information system: John Wheeler's world revisited. *NeuroQuantology*, **2015**, *13*, 57-78.

[306] Jaeger, G. On Wheeler's quantum circuit. In *The quantum-like revolution: A festschrift for Andrei Khrennikov*. Springer International Publishing, Cham, Switzerland, 2023, 25-59.

[307] Duncan, O.D. *Notes on social measurement: Historical and critical*. Russell Sage Foundation, New York, NY, USA 1984.

[308] Fisher, W.P. Jr. Stochastic resonance and Rasch measurement. *Rasch Measurement Transactions*, **1992**, *5*, 186-187 [http://www.rasch.org/rmt/rmt54k.htm].

[309] Fisher, W.P. Jr. Bringing human, social, and natural capital to life: Practical consequences and opportunities. *Journal of Applied Measurement*, **2011**, *12*, 49-66.

[310] Fisher, W.P. Jr. A practical approach to modeling complex adaptive flows in psychology and social science. *Procedia Computer Science*, **2017**, *114*, 165-174.

[311] Fisher, W.P. Jr.; Wilson, M. Building a productive trading zone in educational assessment research and practice. *Pensamiento Educativo: Revista de Investigacion Educacional Latinoamericana*, **2015**, *52*, 55-78.

[312] Crisanti, A.; Falcioni, M.; Paladin, G.; Vulpiani, A. Stochastic resonance in deterministic chaotic systems. *Journal of Physics A: Mathematical and General*, **1994**, *27*,L597.

[313] Petrosky, T.; Prigogine, I. Poincaré resonances and the extension of classical dynamics. *Chaos, Solitons & Fractals*, **1996**, *7*, 441-497.

[314] Prigogine, I.; Rice, S.A. Eds. *Resonances, instability, and irreversibility*, Petrosky, T.; Prigogine, I. eds. John Wiley & Sons, New York, NY, USA 2009. Resonances, instability, and irreversibility: The Liouville space extension of quantum mechanics, 99.

[315] Guttman, L. The basis for scalogram analysis. In Stouffer, S.A.; Guttman, L.; Suchman, E.A.; Lazarsfeld, P.F.; Star, S.A.; Clausen, J.A., Eds. *Measurement and prediction (Studies in social psychology in World War II. volume 4)*. Wiley, New York, NY, USA 1950, 60-90.

[316] Loevinger, J. The attenuation paradox in test theory. *Psychological Bulletin*, **1954**, *51*, 493-504.

[317] Engelhard, G. Jr. What is the attenuation paradox? *Rasch Measurement Transactions*, **1993**, *6*, 257 [http://www.rasch.org/rmt/rmt64.htm].

[318] Masters, G.N. Measurement models for ordered response categories. In Langeheine, R.; Rost, J., Eds. *Latent trait and latent class models*. Plenum Press, Institute for Science Education at the University of Kiel, 1988, 11-30.

[319] Lord, F.M. *Some perspectives on 'The attenuation paradox in test theory'*. ETS, Princeton, New Jersey, USA, 1954, 15 pp.

[320] Lord, F.M. *Applications of item response theory to practical testing problems*. Lawrence Erlbaum, Hillsdale, New Jersey, 1980.

[321] Fisher, W.P. Jr. IRT and confusion about Rasch measurement. *Rasch Measurement Transactions*, **2010**, *24*, 1288 [http://www.rasch.org/rmt/rmt242.pdf].

[322] Adams, R.J.; Wilson, M. ; Wang, W.-C. The Multidimensional Random Coefficients Multinomial Logit Model. *Applied Psychological Measurement*, **1997**, *21*, 1-23.

[323] Adams, R.J.; Wilson, M.; Wu, M. Multilevel item response models: An approach to errors in variables regression. *Journal of Educational and Behavioral Statistics*, **1997**, *22*, 47-76.

[324] Allen, D.D.; Wilson, M. Introducing multidimensional item response modeling in health behavior and health education research. *Health Education Research*, **2006**, *21*, 73-i84.

[325] Beretvas, N.; Kamata, A. Part II. Multi-level measurement Rasch models. In Smith, E.V. Jr. ; Smith, R.M., eds., *Rasch measurement: Advanced and specialized applications* . JAM Press, Maple Grove, MN, (Guest editors) 2007,291-470.

[326] Fischer, G.H. The linear logistic test model as an instrument in educational research. *Acta Psychologica*, **1973**, *37*, 359-374.

[327] Linacre, J.M.; Engelhard, G.; Tatum, D.S.; Myford, C.M. Measurement with judges: Many-faceted conjoint measurement. *International Journal of Educational Research*, **1994**, *21*, 569-577.

[328] von Davier, M., Ed., *Multivariate and mixture distribution Rasch models* . Springer, New York, 2006.

[329] Wilson, M.; Gochyyev, P. Having your cake and eating it too: Multiple dimensions and a composite. *Measurement* , **2020**, *151.*

[330] Andrich, D. A latent trait model for items with response dependencies: Implications for test construction and analysis. In Embretson, S., ed. *Test design: Contributions from psychology, education and psychometrics*. Academic Press, New York, NY, USA 1985, 245-273.

[331] Arminger, G.; Clogg, C.C.; Cheng, T. Regression analysis of multivariate binary response variables using Rasch-type models and finite-mixture methods. *Sociological Methodology*, **2000**, *30*, 1-26.

[332] Eid, M.; Rauber, M. Detecting measurement invariance in organizational surveys. *European Journal of Psychological Assessment*, **2000**, *16*, 20-30.

[333] Haladyna, T.M. Context-dependent item sets. *Educational Measurement*, **1992**, *11*, 21-25.

[334] Hoskens, M.; De Boeck, P. A parametric model for local dependence among test items. *Psychological Methods*, **1997**, *2*, 261-277.

[335] Wilson, M.; Adams, R. Rasch models for item bundles. *Psychometrika*, **1995**, *60*, 181-198.

[336] Wilson, M.; Hoskens, M. The rater bundle model. *Journal of Educational and Behavioral Statistics*, **2001**, *26*, 283-306.

[337] Hendriks, J.; Fyfe, S.; Styles, I.; Skinner, R.; Merriman, G. Scale construction utilising the Rasch unidimensional measurement model: A measurement of adolescent attitudes towards abortion. *Australasian Medical Journal*, **2012**, *5*, 251-261.

[338] Adams, R.J.; Wu, M.; Wilson, M. The Rasch rating model and the disordered threshold controversy. *Educational and Psychological Measurement*, **2012**, *72*, 547-573 [DOI: 10.1177/0013164411432166].

[339] Andrich, D. An expanded derivation of the threshold structure of the polytomous Rasch model that dispels any 'threshold disorder controversy'. *Educational and Psychological Measurement*, **2013**, *73*, 78-124 [DOI: 10.1177/0013164412450877].

[340] Linacre, J.M. Investigating rating scale category utility. *Journal of Outcome Measurement*, **1999**, *3*, 103-122.

[341] Luo, G. The relationship between the rating scale and partial credit models and the implication of disordered thresholds of the Rasch models for polytomous responses. *Journal of Applied Measurement*, **2005**, *6*, 443-55.

[342] Wright, B.D.; Linacre, J.M. Combining and splitting categories. *Rasch Measurement Transactions*, **1992**, *6*, 233-5 [http://www.rasch.org/rmt/rmt63f.htm].

[343] Andrich, D. Models for measurement: Precision and the non-dichotomization of graded responses. *Psychometrika*, **1995**, *60*, 7-26.

[344] Andrich, D.; Marais, I. *A course in Rasch measurement theory: Measuring in the educational, social, and health sciences*. Springer, Cham, Switzerland, Switzerland, 2019.

[345] Andrich, D.; Wright, B.D. Rasch sensitivity and Thurstone insensitivity to graded responses. *Rasch Measurement Transactions*, **1994**, *8*, 382 [http://www.rasch.org/rmt/rmt83m.htm].

[346] Samejima, F. Estimation of latent ability using a response pattern of graded scores. *Psychornetrika Monograph*, **1969**, *17*.

[347] Andrich, D.; Pedler, P. A law of ordinal random error: The Rasch measurement model and random error distributions of ordinal assessments. *Measurement*, **2019**, *131*, 771-781.

[348] Andrich, D.; Surla, D. Equating measuring instruments in the social sciences: Applying measurement principles of the natural sciences. In Fisher, W.P. Jr.; Cano, S., Eds. *Person-centred outcome metrology: Principles and applications for high stakes decision making (Springer Series in Measurement Science & Technology)*. Springer, Cham, Switzerland, 2023, 195-226.

[349] Masters, G.N. Common-person equating with the Rasch model. *Applied Psychological Measurement*, **1985**, *9*, 73-82.

[350] Wolfe, E.W. Understanding Rasch measurement: Equating and item banking with the Rasch model. *Journal of Applied Measurement*, **2000**, *1*, 409-434.

[351] Lord, F.M. The standard error of equipercentile equating. *Journal of Educational Statistics*, **1982**, *7*, 165-174.

[352] Prigogine, I. Time, structure and fluctuations. *Science*, **1978**, *201*, 777-785.

[353] Pattee, H.H. Universal principles of measurement and language functions in evolving systems. In Casti, J.L.; Karlqvist, A. eds. *Complexity, language, and life: Mathematical approaches*. Springer Verlag, Berlin, 1985, 268-281.

[354] Velozo, C.A.; Kielhofner, G.; Lai, J.S. The use of Rasch analysis to produce scale-free measurement of functional ability. *American Journal of Occupational Therapy*, **1999**, *53*, 83-90.

[355] Vass, J. On the Geometry of IFS Fractals and its Applications. [Ph.D. thesis] *diss*. University of Waterloo, Ontario, Canada, Applied Mathematics, 2013.

[356] Peirce, C.S. *The essential Peirce: Selected philosophical writings, volume I (1867-1893)*, Houser, N.; Kloesel, C., Eds. Indiana University Press, Bloomington, 1992.

[357] Sebeok, T.A. In what sense is language a 'primary modeling system'. In Heyvaert, F.J.; Steurs, F., Eds. *World behind words: Essays in honour of Prof. Dr. F. G. Droste on the occasion of his sixtieth birthday*. University of Leuven Press, Leuven, Belgium, 1989, 25-36.

[358] Sebeok, T.A.; Danesi, M. *The forms of meaning: Modeling systems theory and semiotic analysis*. Mouton de Gruyter, Berlin, 2000.

[359] Harvey, P.; Jensen, C.B.; Morita, A. *Infrastructures and social complexity: A companion*. Taylor & Francis.

[360] Lenoir, T. *Inscribing science: Scientific texts and the materiality of communication*. Stanford University Press, Stanford, California, 1998.

[361] Nersessian, N.J. Model-based reasoning in distributed cognitive systems. *Philosophy of Science*, **2006**, *73*, 699-709.

[362] Nersessian, N.J.; Kurz-Milcke, E.; Newstetter, W.C.; Davies, J. Research laboratories as evolving distributed cognitive systems. *Proceedings of the Annual Meeting of the Cognitive Science Society*, **2003**, *25*, 857-862.

[363] Confrey, J.; Shah, M.; Toutkoushian, E. Validation of a learning trajectory-based diagnostic mathematics assessment system as a trading zone. *Frontiers in Education: Assessment, Testing and Applied Measurement*, **2021**, *6*.

[364] Fisher, W.P. Jr. The mathematical metaphysics of measurement and metrology: Towards meaningful quantification in the human sciences. In Morales, A., ed. *Renascent pragmatism: Studies in law and social science*. Ashgate Publishing Co., Brookfield, VT, USA, 2003, 118-153.

[365] Fisher, W.P. Jr. Mathematics, measurement, metaphor, metaphysics: Parts I & II. *Theory & Psychology*, **2003**, *13*, 753-828.

[366] Fisher, W.P. Jr. Meaning and method in the social sciences. *Human Studies: A Journal for Philosophy and the Social Sciences*, **2004**, *27*, 429-454.

[367] Fisher, W.P. Jr. How "deterministic chaos" defines natural laws applicable to both nature and culture. In Durakbasa, N.; Gençyilmaz, M.G., Eds., *Lecture Notes in Mechanical Engineering: Selected papers from ISPR2023, October 5-7, Turkiye*. Springer Verlag, Berlin, 2023.

[368] Lehrer, R. A learning progression emerges in a trading zone of professional community and identity. *WISDOMe Monographs*, **2013**, *3*, 173-186.

[369] Commons, M.L.; Gane-McCalla, R.; Barker, C.D.; Li, E.Y. The model of hierarchical complexity as a measurement system. *Behavioral Development Bulletin*, **2014**, *19*, 9-14.

[370] Jeon, M.; Draney, K.; Wilson, M. A general Saltus LLTM-R for cognitive assessments. In Millsap, R.E.; Bolt, D.M.; van der Ark, L.A.; Wang, W.-C., Eds. *Quantitative Psychology Research. Springer Proceedings in Mathematics & Statistics*. Springer International Publishing, Berlin, 2015. The 78th Annual Meeting of the Psychometric Society, 89, 73-90.

[371] Falmagne, J.-C.; Doignon, J.-P. Stochastic evolution of rationality. *Theory and Decision*, **1997**, *43*, 107-138.

[372] Mislevy, R.J.; & Wilson, M. Marginal maximum likelihood estimation for a psychometric model of discontinuous development. *Psychometrika*, **1996**, *61*, 41-71.

[373] Commons, M.L. Introduction: Attaining a new stage. *Journal of Adult Development*, **2002**, *9*, 155-157.

[374] Ireland, V.; Statsenko, L. Managing complex projects and systems: A literature synthesis. *Australian Journal of Multi-Disciplinary Engineering*, **2020**, *16*, 93-110.

[375] Shannon, C.E. A mathematical theory of communication. In Shannon, C.E.; Weaver, W., Eds. *The mathematical theory of communication*. University of Illinois Press, Urbana, IL, 1949, 29-125.

[376] Layzer, D. The arrow of time. *Scientific American*, **1975**, *233*, 56-69.

[377] Bowman, J.P.; Targowski, A.S., Modeling the communication process: The map is not the territory. *The Journal of Business Communication*, **1987**, *24*, 21-34.

[378] Yuan, X. A symbolic interactionist model of interpreter-facilitated communication-Key communication issues in face-to-face interpreting. *Frontiers in Communication*, **2022**, *7*, 1000849.

[379] Weaver, W. Recent contributions to the mathematical theory of communication. In Shannon, C.E.; Weaver, W., Eds. *The mathematical theory of communication*. University of Illinois Press, Urbana, IL, 1949, 1-28.

[380] Ross, S.N. Fractal transition steps to fractal stages: The dynamics of evolution II. *World Futures: The Journal of New Paradigm Research*, **2008**, *64*, 361-374.

[381] Ross, S.N. Fractal model of nonlinear hierarchical complexity: Measuring transition dynamics as fractals of themselves. *Behavioral Development Bulletin*, **2014**, *19*, 28-.

[382] Badia, X.; Prieto, L.; Roset, M.; Díez-Pírez, A.; Herdman, M. Development of a short osteoporosis quality of life questionnaire by equating items from two existing instruments. *Journal of Clinical Epidemiology*, **2002**, *55*, 32-40.

[383] Dawson, T.L. A comparison of three developmental stage scoring systems. *Journal of Applied Measurement*, **2002**, *3*, 146-89.

[384] Dawson, T.L. New tools, new insights: Kohlberg's moral reasoning stages revisited. *International Journal of Behavioral Development*, **2002**, *26*, 154-66.

[385] Dawson, T.L. A stage is a stage is a stage: A direct comparison of two scoring systems. *Journal of Genetic Psychology*, **2003**, *164*, 335-364.

[386] Dawson, T.L. Assessing intellectual development: Three approaches, one sequence. *Journal of Adult Development*, **2004**, *11*, 71-85.

[387] Fisher, W.P. Jr.; Harvey, R.F.; Kilgore, K.M. New developments in functional assessment: Probabilistic models for gold standards. *NeuroRehabilitation*, **1995**, *5*, 3-25.

[388] Fisher, W.P. Jr.; Harvey, R.F.; Taylor, P.; Kilgore, K.M.; Kelly, C.K. Rehabits: A common language of functional assessment. *Archives of Physical Medicine and Rehabilitation*, **1995**, *76*, 113-122.

[389] Fisher, W.P. Jr.; Eubanks, R.L.; Marier, R.L. Equating the MOS SF36 and the LSU HSI physical functioning scales. *Journal of Outcome Measurement*, **1997**, *1*, 329-362.

[390] Smith, R.M.; Taylor, P. Equating rehabilitation outcome scales: Developing common metrics. *Journal of Applied Measurement*, **2004**, *5*, 229-242.

[391] Bateson, G. Number is different from quantity. *CoEvolution Quarterly*, **1978**, *17*, 44-46 [Reprinted from pp. 53-58 in Bateson, G. (1979). Mind and Nature: A Necessary Unity. New York: E. P. Dutton.].

[392] Fisher, W.P. Jr. Bateson and Wright on number and quantity: How to not separate thinking from its relational context. *Symmetry*, **2021**, *13*.

[393] Wright, B.D. Measuring and counting. *Rasch Measurement Transactions*, **1994**, *8*, 371 [http://www.rasch.org/rmt/rmt83c.htm].

[394] Alker, H.R. A typology of ecological fallacies. In Dogan, M.; Rokkan, S., Eds. *Quantitative ecological analysis in the social sciences*. MIT Press, Cambridge, Massachusetts, 1969, 69-86.

[395] Cash, D.W.; Adger, W.N.; Berkes, F.; Garden, P.; Lebel, L.; Olsson, P.; ..; Young, O. Scale and cross-scale dynamics: Governance and information in a multilevel world. *Ecology and Society*, **2006**, *11*.

[396] Hayman, J.; Rayder, N.; Stenner, A.J.; Madey, D.L. On aggregation, generalization, and utility in educational evaluation. *Educational Evaluation and Policy Analysis*, **1979**, *1*, 31-39.

[397] Kiser, L.L.; Ostrom, E. The three worlds of action: A metatheoretical synthesis of institutional approaches. In Ostrom, E., ed. *Strategies of political inquiry*. Sage, Beverly Hills, 1982, 179-222.

[398] Paul, A.; Lang, J.W.; Baumgartner, R.J. A multilevel approach for assessing business strategies on climate change. *Journal of Cleaner Production*, **2017**, *160*, 50-70.

[399] Rousseau, D.M. Issues of level in organizational research: Multi-level and cross-level perspectives. *Research in Organizational Behavior*, **1985**, *7*, 1-37.

[400] Fisher, W.P. Jr. A nondualist social ethic: Fusing subject and object horizons in measurement. *TMQ--Techniques, Methodologies, and Quality*, **2019**, *10*, 21-40.

[401] Fisher, W.P. Jr.; Pendrill, L.; Lips da Cruz, A.; Felin, A. Why metrology? Fair dealing and efficient markets for the United Nations' Sustainable Development Goals. *Journal of Physics: Conference Series*, **2019**, *1379*.

[402] Fisher, W.P. Jr.; Wilson, M. The BEAR Assessment System Software as a platform for developing and applying UN SDG metrics. *Journal of Physics Conference Series*, **2019**, *1379*.

[403] Fisher, W.P. Jr.; Melin, J.; Möller, C. *Metrology for climate-neutral cities*. RISE Research Institutes of Sweden AB. RISE, Gothenburg, Sweden, 2021, 79 pp.

[404] Fisher, W.P. Jr.; Melin, J.; Möller, C. A preliminary report on metrology for climate-neutral cities. *Acta IMEKO*, **2024**, in press.

[405] Star, S.L.; Ruhleder, K. Steps toward an ecology of infrastructure: Design and access for large information spaces. *Information Systems Research*, **1996**, *7*, 111-134.

[406] Ostrom, E. *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press, Cambridge, UK, 2015. 1990.

[407] Dewey, J. Democracy in education. *The Elementary School Teacher*, **1903**, *4*, 193-204.

[408] Bunderson, C.V.; Newby, V.A. The relationships among design experiments, invariant measurement scales, and domain theories. *Journal of Applied Measurement*, **2009**, *10*, 117-137.

[409] De Boeck, P.; Wilson, M. Eds. *Explanatory item response models: A generalized linear and nonlinear approach*. Springer-Verlag, New York, NY, USA 2004. Statistics for Social and Behavioral Sciences.

[410] Embretson, S.E. *Measuring psychological constructs: Advances in model-based approaches*. American Psychological Association, Washington, DC, 2010.

[411] Graßhoff, U.; Holling, H.; Schwabe, R. Optimal designs for linear logistic test models. In Giovagnoli, A.; Atkinson, A.C.; Torsney, B.; May, C., Eds. *MODa9 -- Advances in Model-Oriented Design and Analysis: Contributions to statistics*. Physica-Verlag HD, Heidelberg, Germany, 2010, 97-104.

[412] Green, K.E.; Kluever, R.C. Components of item difficulty of Raven's Matrices. *Journal of General Psychology*, **1992**, *119*, 189-199.

[413] Green, K.E.; Smith, R.M. A comparison of two methods of decomposing item difficulties. *Journal of Educational Statistics*, **1987**, *12*, 369-381.

[414] Morell, L.; Collier, T.; Black, P.; Wilson, M. A construct-modeling approach to develop a learning progression of how students understand the structure of matter. *Journal of Research in Science Teaching*, **2017**, *54*, 1024-1048.

[415] Pandarova, I.; Schmidt, T.; Hartig, J.; Boubekki, A.; Jones, R.D.; Brefeld, U. Predicting the difficulty of exercise items for dynamic difficulty adaptation in adaptive language tutoring. *International Journal of Artificial Intelligence in Education*, **2019**, *29*, 342-367.

[416] Smith, R.M. Item component equating. In Wilson, M., ed. *Objective measurement: Theory into practice, Vol. 3*. Ablex Publishing Co., Norwood, New Jersey, 1996, 289-308.

[417] Sonnleitner, P. Using the LLTM to evaluate an item-generating system for reading comprehension. *Psychology Science Quarterly*, **2008**, *50*, 345-362.

[418] Stenner, A.J.; Smith, M. III; Burdick, D.S. Toward a theory of construct definition. *Journal of Educational Measurement*, **1983**, *20*, 305-316.

[419] Stenner, A.J.; Fisher, W.P. Jr.; Stone, M.H.; Burdick, D.S. Causal Rasch models. *Frontiers in Psychology: Quantitative Psychology and Measurement*, **2013**, *4*, 1-14.

[420] Chan, T.L.; Perlmutter, M.S.; Andrews, M.; Sunness, J.S.; Goldstein, J.E.; Massof, R.W.; Low Vision Research Network (LOVRNET) Study Group, Equating visual function scales to facilitate reporting of Medicare functional g-code severity/complexity modifiers for low-vision patients. *Archives of Physical Medicine and Rehabilitation*, **2015**, *96*, 1859-1865.

[421] Hobart, J.C.; Cano, S.J.; Zajicek, J.P.; Thompson, A.J. Rating scales as outcome measures for clinical trials in neurology: Problems, solutions, and recommendations. *Lancet Neurology*, **2007**, *6*, 1094-1105.

[422] Cano, S.; Klassen, A.F.; Scott, A.; Thoma, A.; Feeny, D.; Pusic, A. Health outcome and economic measurement in breast cancer surgery: Challenges and opportunities. *Expert Review of Pharmacoeconomics & Outcomes Research*, **2010**, *10*, 583-594.

[423] Fisher, W.P. Jr.; Oon, E.P.-T.; Benson, S. Rethinking the role of educational assessment in classroom communities: How can design thinking address the problems of coherence and complexity? *Educational Design Research*, **2021**, *5*, 1-33.

[424] Draney, K.; Wilson, M. Selecting cut scores with a composite of item types: The construct mapping procedure. *Journal of Applied Measurement*, **2011**, *12*, 298-309.

[425] Wyse, A.E. Construct maps as a foundation for standard setting. *Measurement: Interdisciplinary Research and Perspectives*, **2013**, *11*, 139-170.

[426] Chien, T.W.; Chang, Y.; Wen, K.S.; Uen, Y.H. Using graphical representations to enhance the quality-of-care for colorectal cancer patients. *European Journal of Cancer Care*, **2018**, *27*,e12591.

[427] Mead, R.J. The ISR: Intelligent Student Reports. *Journal of Applied Measurement*, **2009**, *10*, 208-224.

[428] Wright, B.D.; Mead, R.; Bell, S.R. *BICAL: Calibrating items with the Rasch model*. University of Chicago, Illinois, USA, Statistical Laboratory, Department of Education, Chicago, Illinois, USA, 1980, 170 pp.

[429] Hertz, T.; Schlüter, M. The SES-framework as boundary object to address theory orientation in social-ecological system research: The SES-TheOr approach. *Ecological Economics*, **2015**, *116*, 12-24.

[430] Karasti, H. Infrastructuring in participatory design. *Proceedings of the 13th Participatory Design Conference: Research Papers*, **2014**, *1*, 141-150.

[431] Karasti, H.; Millerand, F.; Hine, C.M.; Bowker, G.C. Knowledge infrastructures: Intro to Part I. *Science & Technology Studies*, **2016**, *29*, 2-12.

[432] Halsema, A. The subject of critique: Ricoeur in dialogue with feminist philosophers. *Etudes Ricoeuriennes/Ricoeur Studies*, **2013**, *4*, 21-39.

[433] Ricoeur, P. The critique of subjectivity and cogito in the philosophy of Heidegger. In Frings, M.S., ed. *Heidegger and the quest for truth*. Quadrangle Books, Chicago, Illinois, USA, 1968, 62-75.

[434] Bloom, B.S. The hands and feet of genius: Automaticity. *Educational Leadership*, **1986**, *43*, 70-77.

[435] Cooper, S.; Hebert, M.; Goodrich, J. M.; Leiva, S.; Lin, X.; Peng, P.; Nelson, J. R. Effects of automaticity training on reading performance: A meta-analysis. *Journal of Behavioral Education*, **2022**, 1-30.

[436] Kahneman, D. *Thinking fast and slow*. Farrar, Straus and Giroux, New York, NY, USA 2011.

[437] Rose, T. *Collective illusion: Conformity, complicity, and the science of why we make bad decisions*. Hachette, New York, NY, USA 2022.

[438] Nadler, S.; Shapiro, L. *When bad thinking happens to good people: How philosophy can save us from ourselves*. Princeton University Press, Princeton, New Jersey, USA, 2021.

[439] Stengers, I. *Another science is possible: A manifesto for slow science.* Polity Press, Cambridge, UK, 2018.

[440] Hayek, F.A. *Individualism and economic order*. University of Chicago Press, Chicago, Illinois, USA, 1948.

[441] Whitehead, A.N. *An introduction to mathematics*. Henry Holt and Co., New York, NY, USA 1911.

[442] Ihde, D. *Bodies in technology*. University of Minnesota Press, Minneapolis, Minnesota, USA, 2002.

[443] Overton, W.F. Embodiment from a relational perspective. In Overton, W.F.; Muller, U.; Newman, J.L., Eds. *Developmental perspective on embodiment and consciousness*. Erlbaum, Mahwah, New Jersey, USA, 2008, 1-18.

[444] Fisher, C.M.; Demir-Caliskan, O.; Yingying Hua, M.; A. Cronin, M. Trying not to try: The paradox of intentionality in jazz improvisation and its implications for organizational scholarship. In Bednarek, R.; Pina e Cunha, M.; Schad, J.; Smith, W.K., Eds. *Research in the Sociology of Organizations*, Volume 73b *Interdisciplinary Dialogues on Organizational Paradox: Investigating Social Structures and Human Expression, Part B*. Emerald Publishing Limited, Bingley, 2021, 123-137.

[445] Abram, D. *The spell of the sensuous: Perception and language in a more-than-human world*. Vintage Books, New York, NY, USA 1996.

[446] Stengers, I. *The invention of modern science*, Smith, D.W., trans. University of Minnesota Press, Minneapolis, Minnesota, USA, 2000. Theory out of bounds, 19.

[447] de Solla Price, D.J. Of sealing wax and string. In *Little science, big science--and beyond*. Columbia University Press, New York, NY, USA 1986, 237-253.

[448] Babich, B.E. Material hermeneutics and Heelan's philosophy of technoscience. *AI & Society*, **2020**, 1-12.

[449] Ihde, D. The historical and ontological priority of technology over science. In *Existential technics*. State University of New York Press, Albany, New York, NY, USA 1983, 25-46.

[450] Foxon, T. Bounded rationality and hierarchical complexity: Two paths from Simon to ecological and evolutionary economics. *Complexity and Ecological Economics*, **2006**, *3*, 361-368.

[451] Kahneman, D. Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review*, **2003**, *93*, 1449-1475.

[452] Wittgenstein, L. *Tractatus logico-philosophicus*. Harcourt Brace, London, UK, 1922.

[453] Martland, T.R. On "The limits of my language mean the limits of my world". *The Review of Metaphysics*, **1975**, *29*, 19-26.

[454] Mach, E. *The science of mechanics: A critical and historical account of its development*, McCormack, T.J., trans. The Open Court Publishing Co., Chicago, Illinois, USA, 1919.

[455] Banks, E. The philosophical roots of Ernst Mach's economy of thought. *Synthese*, **2004**, *139*, 23-53.

[456] Franck, G. The scientific economy of attention: A novel approach to the collective rationality of science. *Scientometrics*, **2002**, *55*, 3-26.

[457] Franck, G. The economy of attention. *Journal of Sociology*, **2019**, *55*, 8-19.

[458] Heidegger, M. *What is called thinking?* Gray, J.G., trans. Harper & Row, New York, NY, USA 1968.

[459] North, D.C. *Structure and change in economic history*. W. W. Norton & Co., New York, NY, USA 1981.

[460] Wittgenstein, L. 1953 *Philosophical investigations*, Anscombe, G.E.M., trans. Macmillan, New York, NY, USA 1958.

[461] Pastor-Satorras, R.; Castellano, C.; Van Mieghem, P.; Vespignani, A. Epidemic processes in complex networks. *Reviews of Modern Physics*, **2015**, *87*, 925.

[462] Pentland, A. *Social physics: How good ideas spread-the lessons from a new science*. Penguin, London, UK, 2014.

[463] Diener, P. The evolution of dysfunction. *Culture*, **1982**, *2*, 43-51.

[464] Moore, R. *Niels Bohr: The man and the scientist*. Hodder & Stoughton, London, 1967.

[465] Sahlins, M. *The new science of the enchanted universe: An anthropology of most of humanity*. Princeton University Press, Princeton, New Jersey, USA, 2022.

[466] Kauffmann, S. *At home in the universe: The search for laws of self-organization and complexity*. Oxford University Press, New York, NY, USA 1996.

[467] Ehrenfest-Afanassjewa, T. Zur Axiomatisierung des Zweiten Hauptsatzes der Thermodynamik [On the axiomatization of the second law of thermodynamics]. *Zeitschrift für Physik*, **1925**, *33*, 933-946.
[468] Ehrenfest-Afanassjewa, T. *Relevia: Een nieuw economisch systeem, een orde, waarin ik zelf ook graag zou willen leven*. [*Relevia: A new economic system, an order in which I myself would like to live.*] Boucher, The Hague, 1946.
[469] Prigogine, I.; Defay, R. *Thermodynamique chimique*. Dunod, Paris, 1944.

International Journal of Advances in Production Research

Machine Learning Methods for (Dis-)Assembly Sequence Planning - A Systematic Literature Review

Detlef Gerhard[a,1*], Julian Rolf[a,2], Pascalis Trentsios[a,3], Jan Luca Siewert[a,4]
[a] *Digital Engineering Chair, Ruhr University Bochum, Germany*
{detlef.gerhard[1], julian.rolf[2], pascalis.trentsios[3], jan.siewert[4]}
@ruhr-uni-bochum.de

Abstract— **This paper presents a systematic literature review on the application of reinforcement learning in the domain of assembly and disassembly sequence planning. The authors conduct a keyword search to identify scientific publica-tions in the desired field in three scientific databases. Web of Science, Scopus and IEEE-Xplore. The analysis covers two core aspects of reinforcement learning, namely the definition of the reward function and the representation of states. In total 23 publications are identified, and the content of the collected works is presented. An analysis of the selected publications is then carried out in relation to the questions posed in order to be able to make recommendations for the application of reinforcement learning methods for the generation of efficient assembly and demonstration se-quences.**

Keywords— **reinforcement learning; assembly sequence planning; disassembly sequence planning**

I. INTRODUCTION

Sequence planning for assembly tasks is a fundamental step in manufacturing and industrial automation, crucial for ensuring efficient and error-free production processes. The time of assembling takes up 20-50 percent of the total time of production, while the cost of assembling takes up about 20-30 percent of the total cost [1,2], while the optimization of the disassembly also offers time and cost savings [3]. Given the fact, that scheduling both an assembly and disassembly sequence can be modelled as a NP-hard problem, those solutions are limited in handling the increasing complexity of modern manufacturing without exceeding computational limitations. Therefore, many different approaches to find semi-optimal solutions have been considered in literature.

Modelling the assembly as a graph allows the use of graph search algorithms to find an optimal or sufficiently good solution, depending on the algorithm [4,5]. Heuristic methods can also be used to solve ASP/ DSP problems. These include genetic algorithms (GA) [6,7]. GAs represent (dis)assembly sequences as chromosomes, which are iteratively adapted by mutations and recombination. Afterwards solutions get selected with regard to an evaluation scheme. This iteration procedure is repeated until a sufficient solution is found. Alternatively, neural networks can be trained to decide for each step which component should be (dis)assembled next [8]. These can be trained using supervised learning methods, which requires the availability of a suitable data set. This requirement does not apply to reinforcement learning and is therefore particularly advantageous in the planning phase of a product when no useful data from a real (dis)assembly is available.

This literature review provides an overview of the state of the research regarding the application of reinforce-ment learning methods for the generation of assembly or disassembly sequences based on a systematic literature review. To this end, the basics of reinforcement learning are presented first. The research questions on which the literature review is based are then defined and the procedure for identifying relevant studies is described. The results of the literature review are then presented and analysed. Finally, recommendations for the definition of rewards and state descriptions in the field of reinforcement learning for the generation of (dis)assembly sequences are given and potential research.

Deep Reinforcement Learning (DRL) falls under the broader category of machine learning, a core component of artificial intelligence (AI). As outlined by Sutton and Barto, DRL involves three primary elements: an entity known as the agent, the environment in which it operates, and a specific objective or task [9]. Within this frame-work, DRL incorporates several critical aspects, including a policy, a reward function, a value function, and some-times a model representing the environment.

In DRL, the agent has the capability to perceive the state of its environment, interact with it through various actions, and receive rewards based on successful task completion. The policy, which determines how actions are chosen in response to the states perceived, is developed through training, where rewards or penalties are given according to the reward function. The value function predicts the total expected reward from a particular action in the current state. In certain DRL approaches, providing the agent with a model of the environment enables more effective action planning.

*Corresponding Author

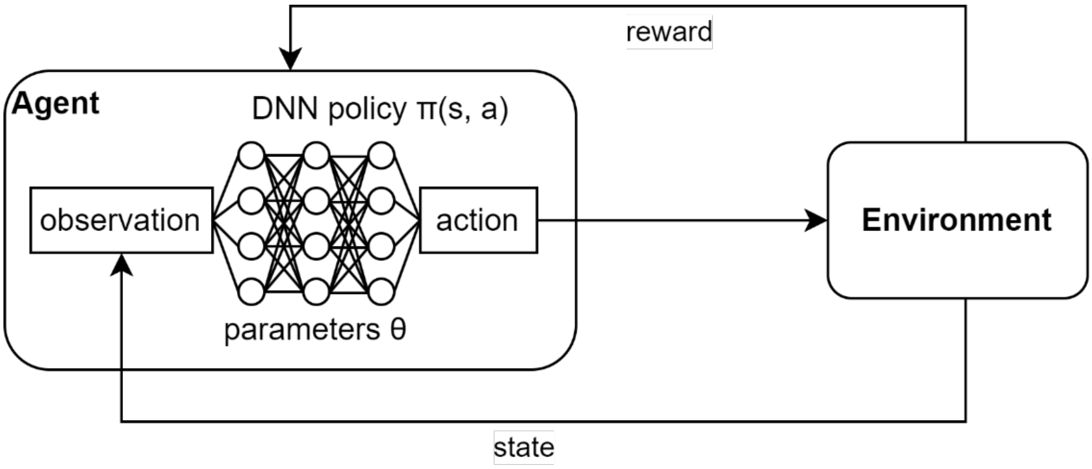


FIGURE 1. Implemented Components for the Assembly Support System

DRL is distinguished from traditional reinforcement learning by its incorporation of deep neural networks (DNNs) to formulate the agent's policy. The versatility of DNNs as universal function approximations renders them effective in various applications [10]. The structure of these networks can vary significantly, depending on both the task at hand and the architecture of the agent or system [11]. The training process in DRL involves repetitive agent-environment interactions, as depicted in Figure 1. Each interaction includes the perception of the current state, execution of an action, and receipt of a reward. The state encapsulates comprehensive information about the system, environment, and their ongoing interaction [12]. Training episodes begin at a start state and conclude upon meeting specific conditions, like successful task completion, failure, or reaching a predetermined time or step limit. These episodes and individual steps generate data used to adjust the DNN's parameters or the policy, aiming to maximize the agent's total reward. Training data should be diverse to effectively prepare for real-world application post-training. Randomizing the initial state in each episode contributes to this diversity. Training effectiveness is also influenced by various hyperparameters [13]. Policy updates depend on the application of rewards through the reward function, which directs the system's behaviour by assessing the states and interactions among the agent and its environment. These interactions are classified as either beneficial or not. The reward function sets the parameters for desired system behaviour, operating independently of the system's state or action space and not directly influencing its sensors or actuators [14].

II. MATERIALS AND METHODS

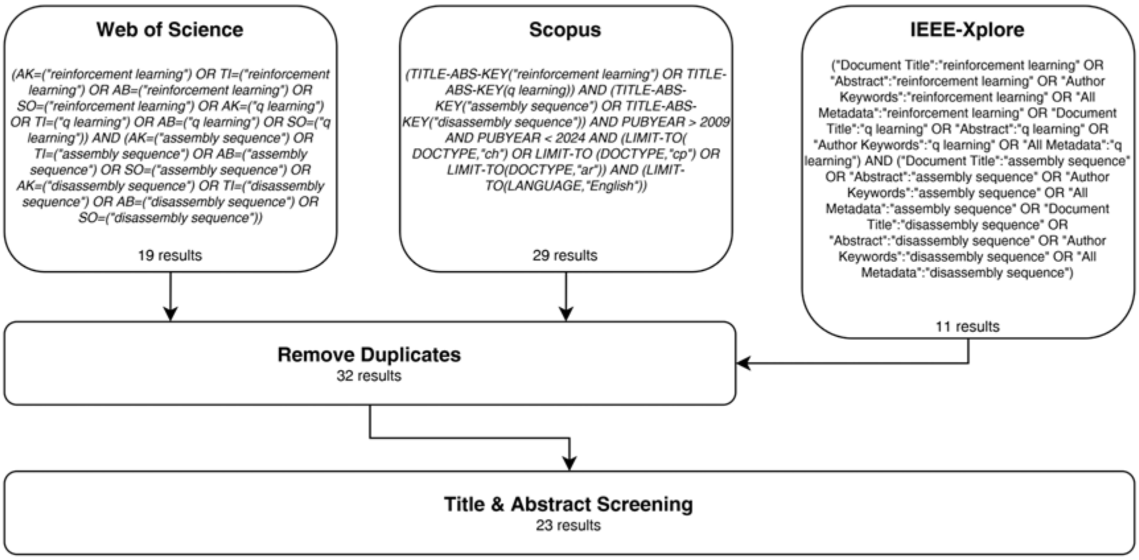
This section provides information about the research questions defined to analyse the scientific works. Additionally, the procedure for identifying relevant scientific work is presented.

A. Research Questions

The scientific analysis focuses on two research questions, aimed at illustrating key aspects for applying reinforcement learning in the context of assembly and disassembly sequence planning. The main difficulties lie particularly in the description of the state and the reward for the agent’s actions. The state must contain all relevant information that influences the agent’s decision. At the same time, high dimensional input requires more data samples to identify underlying correlations. This leads to an increased training effort in RL. Therefore, it is mandatory to choose a state description that is both complete and sufficiently extensive.

The reward leads the agent to the correct policy. Accordingly, a suitable value must be assigned to all good and bad actions to be assigned. The definition of a reward function is often associated with a trial-and-error procedure. Thus, the systematic literature review provides a starting point for defining the reward function to simplify the RL application. The research questions are as follows:
RQ 1: Which factors should be considered when defining an agent's reward function for reinforcement learning, and which conditions should be fulfilled to do so?
RQ 2: Which state representation for the reinforcement learning agent should be chosen and under what conditions?

B. Keyword Analysis



```
(AK=("reinforcement learning") OR TI=("reinforcement learning") OR AB=("reinforcement learning") OR SO=("reinforcement learning") OR AK=("q learning") OR TI=("q learning") OR AB=("q learning") OR SO=("q learning")) AND (AK=("assembly sequence") OR TI=("assembly sequence") OR AB=("assembly sequence") OR SO=("assembly sequence") OR AK=("disassembly sequence") OR TI=("disassembly sequence") OR AB=("disassembly sequence") OR SO=("disassembly sequence"))
```

```
(TITLE-ABS-KEY("reinforcement learning") OR TITLE-ABS-KEY(q learning)) AND (TITLE-ABS-KEY("assembly sequence") OR TITLE-ABS-KEY("disassembly sequence")) AND PUBYEAR > 2009 AND PUBYEAR < 2024 AND (LIMIT-TO( DOCTYPE,"ch") OR LIMIT-TO (DOCTYPE,"cp") OR LIMIT-TO(DOCTYPE,"ar")) AND (LIMIT-TO(LANGUAGE,"English"))
```

```
("Document Title":"reinforcement learning" OR "Abstract":"reinforcement learning" OR "Author Keywords":"reinforcement learning" OR "All Metadata":"reinforcement learning" OR "Document Title":"q learning" OR "Abstract":"q learning" OR "Author Keywords":"q learning" OR "All Metadata":"q learning") AND ("Document Title":"assembly sequence" OR "Abstract":"assembly sequence" OR "Author Keywords":"assembly sequence" OR "All Metadata":"assembly sequence" OR "Document Title":"disassembly sequence" OR "Abstract":"disassembly sequence" OR "Author Keywords":"disassembly sequence" OR "All Metadata":"disassembly sequence")
```

FIGURE 2. Process of the systematic literature search

The workflow for identifying relevant scientific papers for this structural review analysis involved a systematic keyword analysis across three reputable databases: Scopus, IEEE-Xplore, and Web of Science (WoS). The first step in this process is to define the research scope, and to establish the parameters that would guide the subsequent literature search. The search terms chosen are “reinforcement learning” or “q learning” to determine the methodology, combined with “assembly sequence” or “disassembly sequence” to define the field of application. Publications released before 2010 were not considered. Scientific papers are identified by the existence of the keyword combinations in the authors keywords, abstracts, or titles, ensuring a comprehensive representation of the targeted domain. The search in the IEEE-Xplore database also includes the metadata field, while WoS search also includes the "plus keywords".

After performing the search, publications are filtered for duplicates using their DOI. The remaining works are filtered by screening the title and abstract, checking their potential for answering the defined research questions. The process of the systematic literature search is visualized in figure 2.

III. RESULTS

This section presents the scientific papers identified. These are categorized according to their focus on assembly sequence planning or disassembly sequence planning.

In the total, 23 scientific papers were identified that match the described criteria They span a publication period from 2013 to 2023, with 22 papers released in the last 4 years. This demonstrates the actuality of the topic as described in Figure 3.

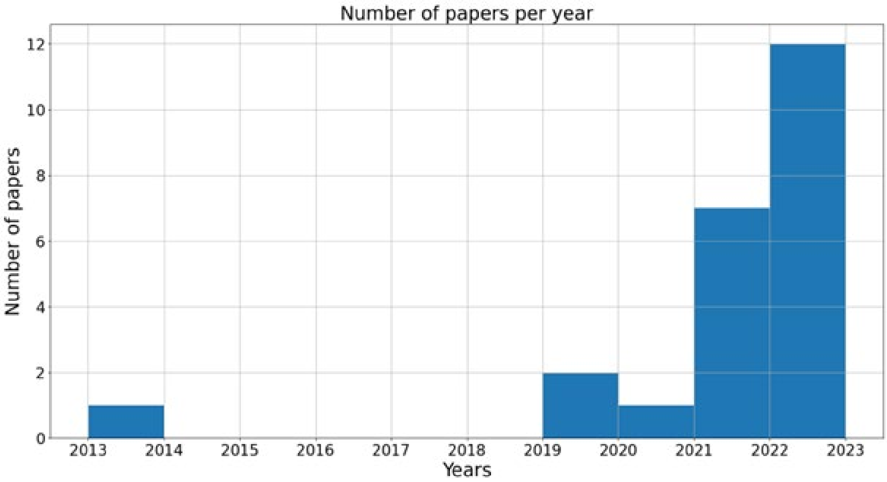


FIGURE 3. Annual Releases

An analysis of the collaboration network in Figure 4 revealed three dominant author clusters with multiple publications to the domain of applying reinforcement learning for assembly or disassembly sequence planning using bibliometrix [15].

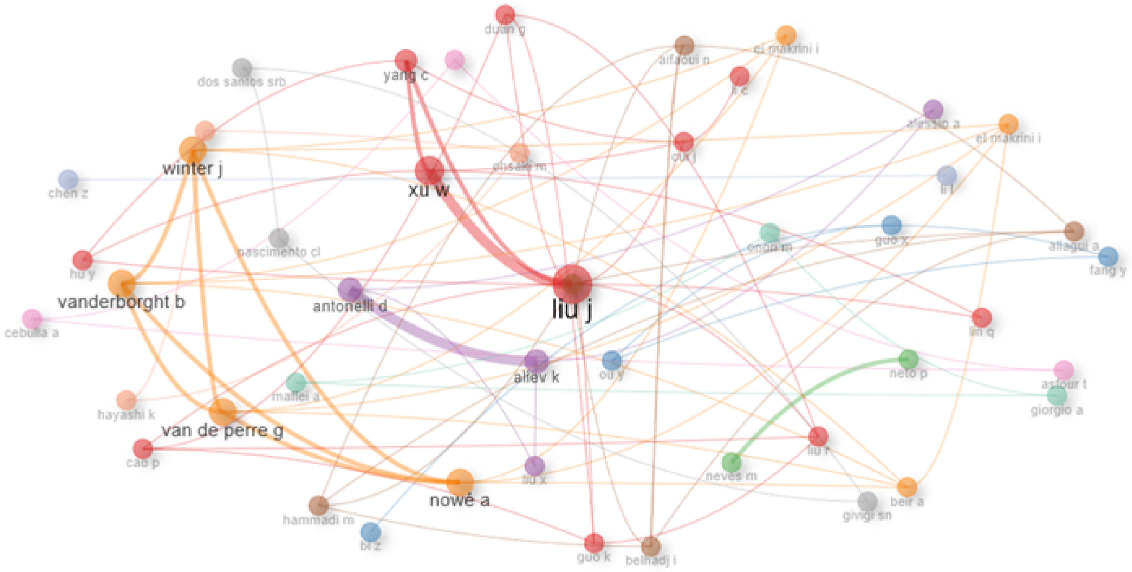


FIGURE 4. Collaboration Network

An overview of the selected works is given in Table 1. For each work, the selection of the algorithm used, the chosen approach, the properties to be optimized, the state description and the number of components in the assembly being studied are given.

Table1 1
Overview of the selected works

Publication	Algorithm	Approach	Objective	State Desc.	Assembly Parts
Neves [16]	Q-Learning	Assembly	Duration	Binary List + Selected Tool	9
Neves [17]	Tabular Q-Learning, A2C, Deep Q-Learning, Rainbow	Assembly	Duration, Assembler Feedback	Binary List + Selected Tool	9
Giorgio [18]	Online RL	Assembly	Duration, Assembler Feedback	-	21

Hayashi [19]	RL, GA	Assembly by Disassembly	Number of temporary support elements	Connectivity Matrix	-
Kitz [20]	Q-Learning	Assembly by Disassembly	Collision free Space	2 ½ D Collision Matrix	7, 16, 56
Zhao [21]	Deep Q-Learning	Assembly	Correctly Assembled	Part & Assembly Image	7
Antonelli [22]	Q-Learning	Assembly Human + Robot	Number Steps, Reaching Final States	Discrete Number	10
Antonelli [23]	RL	Assembly Human + Robot	Reaching Final States	Discrete Number	10
Yin [24]	RL	Robotic Assembly	Assembly Time, Energy Consumption	Connectivity Matrix	8
Alessio [25]	Multi Agent RL	Assembly Human + Robot	Reaching Final States	Grid-Layout	10
Dos Santos [26]	Learning Automata	Robotic Assembly	Robotic Movement Adaptions	Grid-Layout	3 Floor Tower
Watanabe [27]	Q-Learning	Robotic Assembly, Assembly by Disassembly	Duration	Binary List	-
Winter [28]	Hierarchical RL	Assembly, Human + Robot	Number Steps	Binary List	Cranfield Benchmark
Winter [29]	Q-Learning	Robotic Assembly	Duration	Binary List	Cranfield Benchmark
Cebulla [30]	Monte Carlo Tree Search, Q-Learning	Assembly by Disassembly	Accessibility, Directional Changes	2 ½ D Collision Matrix	38, 58
Guo [31]	Deep RL	Assembly	Directional Changes	Connector Matrix	20, 30, 40, 50
Bi [32]	Q-Learning	Disassembly	Component wise Reward	AND/OR Graph	6, 7, 22
Chen [33]	Q-Learning	Disassembly	Component wise Reward, Duration	Binary List	11
Yang [34]	Deep RL, Bee Algorithm, GA	Robotic Disassembly	Tool switch, Directional Changes	Binary List	21
Zhao [35]	Deep Q-Learning	Disassembly	Component wise Reward, Duration	MSDHGM	26

Cui [36]	Deep Q-Learning	Robotic Disassembly	Tool switch, Directional Changes	Binary List	21
Allagui [37]	Q-Learning	Disassembly	Duration, Directional Changes	Collision Matrix	5
Liu [38]	Deep Q-Learning	Robotic Disassembly	Duration	Binary List	17, 23

A. Assembly Sequence Planning

Neves et al. [16,17] use the duration of the assembly to determine the reward. To determine the duration, the authors normalize the required time t of an assembly step between the minimum assembly time T_{min} and the maximum assembly time T_{max}. These two bounds are defined by empirical values from previous episodes. To reduce outliers in the value of the maximum assembly time, only values of the assembly time are considered if they lie within 2 standard deviations of the durations of the previous 100 episodes.

$$R_t = \frac{(T_{max} - t)}{(T_{max} - T_{min})}$$

The system also checks whether the tool required for the assembly step needs to be changed. The tool switch time is included in the reward function as well. The object used in a study case is an airplane toy, which consists of 9 structural parts and 2 types of fasteners. The assembly process is divided into 8 different tasks, resulting in a total of 40320 possible sequences, and 3360 feasible ones. The authors conduct multiple different training scenarios, altering the measurement of the task time (averaging over 10 repeated times) or restricting impossible actions. By averaging the task time and without the restriction of impossible actions, the authors state the optimal training parameters with 6500 max. training episodes.

In [17], it is also argued that, in addition to considering the assembly time, feedback from the assembler should also be considered when awarding the reward. Properties to be considered can be the difficulty of the assembly step or ergonomics. The status description of the assembly is described by a list with N_{a+1} elements. The first N_a entries refer to the possible tasks. If a task is fulfilled, the corresponding value in the list is set to 1. Otherwise, it has the value 0. The last additional position describes the selected tool, which is defined by a discrete number, whereby the value 0 stands for no selected tool. The selection of the agent's action is also described by a discrete value.

One focus of the work [17] is the comparison of different RL algorithms. These are: Tabular Q-learning, A2C, DQN and Rainbow. During training, the agent had to find an optimal assembly sequence using the same assembly as in [16]. The training results state, that the DQN algorithm presented the worst performance, reaching suboptimal assembly sequence time durations and experiencing a 4 to 5 times higher number of unwanted assembly sequences when compared to the other 3 algorithms. The algorithms tabular Q-Learning, A2C, and Rainbow had similar performances and were able to achieve near optimal assembly times in both deterministic and stochastic scenarios, after approximately 10.000 episodes.

The authors assume, that the Q-learning algorithm is known to suffer from the “curse of dimensionality”, so that by increasing the assembly complexity, an even worse result is to be expected. Giorgio et al. [18] argue that online RL helps to shorten the feedback path in assembly planning. Traditionally, the assembler can only give feedback on the design or arrangement of the components and assemblies after assembly, whereas with RL additional feedback can be given for each individual assembly step. Although this "fast feedback" only serves to optimize the assembly sequence and not to adapt the design, it can still improve the process in sufficient quantities. The authors consider training in a real environment instead of a virtual one. In the first publication [18], data is collected for training the RL approach. The assembly consists of 21 components, including the connecting elements. The exact number of possible assemblies is not specified. Giorgio et al. argue to divide the reward function into two categories: the satisfaction of the assembler and the numerical characteristics of the assembly, such as the duration. The definition of observations, actions and states is not specified.

Hayashi et al. [19] consider the use case of the assembly of truss systems. The work focuses on linking the RL approach for finding effective assembly sequences and the structural analysis of the assembly. The authors use the assembly by disassembly approach. Starting from the complete state of the assembly, one element of the assembly is removed per step until only one element remains. The agent is responsible for selecting the element to be removed in each step. At the end, the disassembly sequence is reversed and considered as the assembly sequence.

The aim of the authors is to minimize the number of support elements required to ensure the structural integrity of the assembly. The reward is determined accordingly from the number of temporary support elements multiplied by -1. The state of the assembly is represented by a connectivity matrix C and an input matrix $\hat{v}$, which represents the inputs for each node with binary flags to distinguish permanent pin-supports and locally unstable nodes. The agent's action consists of removing a component. For the training, validation and testing of the RL model, truss structures of different sizes are selected. However, the exact number of components is not specified. The training took 2.5h to complete the 5000-episode long training on multiple different spatial trusses. Afterwards, the trained model is validated on three different trusses, to show its general applicability.

Kitz et al. [20] also use the assembly by disassembly approach. Collision checks are carried out to determine the dismantlability of an element. In addition, it is ensured that the stability of the assembly is guaranteed after the disassembly of a component. The stereographic projection of the respective collision-free vectors of each component is stored in a 2 ½ D collision map and describes how far a component can be moved along a vector until a collision occurs. The collision-free space of an element is used to determine the dismantling costs. The greater the space, the lower the costs and the greater the reward for the agent. If the agent chooses an invalid action, it is penalized with a fixed negative reward. The state is described using a vector with binary entries for each element of the assembly. To validate the concept, the authors used three assemblies with 7, 16 and 56 parts. The total training episodes are 1322, 2000 and 3741 and the agent is reliably able to determine cost-minimal sequences.

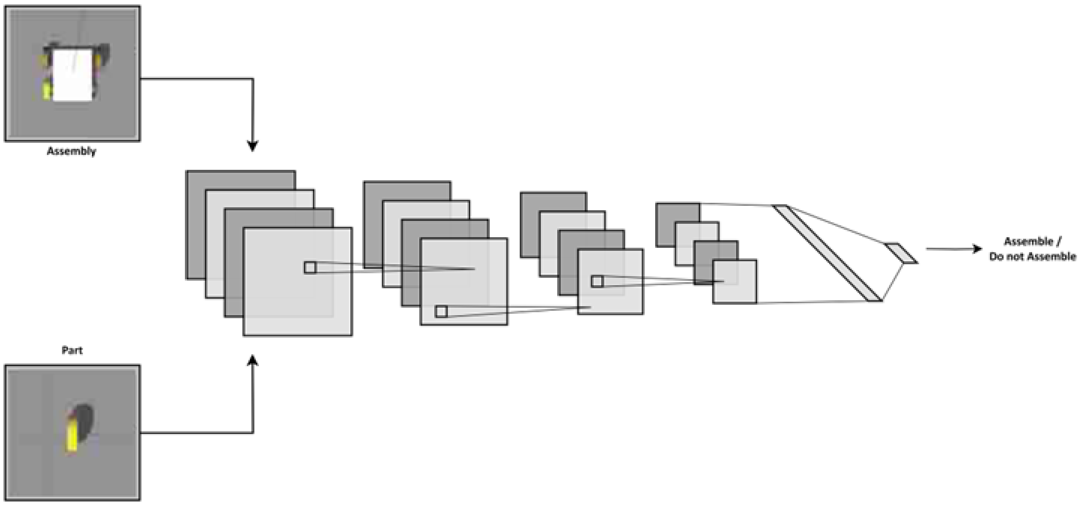


FIGURE 5. Structure of the neural network used by Zhao et al. [21]

Zhao et al. [21] consider the formulation and evaluation of a generally applicable concept for generating assembly sequences of components using RL. In each step, the agent decides for a selected element whether it should be assembled or not. Accordingly, the agent's action is a Boolean value. The description of the status consists of two parts. Firstly, the status of the assembly, represented by a camera image, and secondly, the component whose assembly must be decided, also represented by an image. Both images are feed into a convolutional neural network (CNN), as described in figure 5. The agent receives a positive reward of 1 if all elements of the assembly have been correctly selected from the set of available elements, otherwise, a negative reward of -1 is awarded. To shorten the training time and improve the performance of the CNN used, the authors apply Curriculum Learning and Transfer Learning. The concept is first validated in a virtual environment by assembling an assembly with seven elements. Using curriculum learning, the model can determine a correct assembly sequence in 85% of cases with random initialization of the environment. With parameter transfer, the training time is about twice as fast. To test the general applicability, the authors create further use cases. In each case, around 100 further training episodes are required to generate suitable assembly sequences.

Antonelli et al. [22,23] consider the cooperation between humans and robots. The authors focus on the agile adaptation of the sequence by the human actor in the system. Accordingly, the robot must react to this adaptation by planning a new assembly sequence. First, the assembly steps are grouped according to the actors (executable only by the human/robot, executable both by the human/robot).

In [22], the awarding of rewards is based on the achievement of states. In [23], on the other hand, a reward of -1 is awarded for each action performed. Only the achievement of final configurations leads to a positive reward of 2. A state is described by a discrete number. The case study includes 26 possible Operations to pick from, with 15 different feasible states of the assembly. Assembly is made from a base part, on which three flanges are mounted and joined by screwed bolts. An optimal assembly sequence was found after approx. 350 steps, conducting of 3 assembly operations.

The generation of an optimal assembly sequence for a scenario in which the cooperation of several two-armed assembly robots is necessary is considered by Yin et al. [24]. In this scenario, truss elements must be assembled by robots. These consist of bi-directional tubes and connecting elements at the ends. Each robot makes one of three possible decisions per step:

1. supporting a connecting element
2. the assembly of a pipe
3. adjusting the position

A state is described by four components. A truss matrix H_s. The number of rows and columns of the matrix corresponds to the number of connecting elements. An element in row i and column j describes the state of a pipe between the corresponding connecting elements. The state is described by a Boolean value. The actions currently performed by the robots as a nested list A_s. The current position of each robot is a list of vectors O_s. A list of Boolean values F_s describes whether a robot is currently carrying a pipe. The reward is based on the required assembly time, the energy consumption of the robots and the stability of the setup. Two scenarios, each with two robots, were considered to validate the concept. In the experiment, the sequence of two mobile dual arm robots assembling a triangular pyramid truss structure is optimized, with eight rods. The length of the training comprises around 3000 episodes. However, the reward between the episodes is still very variable towards the end and seems to converge only slightly.

As in [22,23], Alessio et al. [25] also consider the cooperation between a human agent and a robot as a multi-agent reinforcement learning scenario. However, the human agent is designed in such a way that errors occur in its behaviour. The aim of the work is to recognize human errors and to determine suitable reactions to these errors. The virtual environment is designed as a grid layout, with each cell representing possible states of the agents, based on the Markov Decision Process, which can be described using a graph (figure 6). The evaluation of the human agent and the robot is slightly different but follows the same principle. Positive rewards are awarded for achieving defined final states. Non-permissible actions or idling are punished with negative rewards. Any other permissible action is penalized with a reward of -1 to shorten the duration of the assembly. Since the robot agent's behaviour is to be improved as part of the work, 10% of the human agent's reward is deducted from the robot's reward. In this way, possible human misbehaviour is mapped, and the robot is forced to adapt its behaviour accordingly. Four properties are used to describe the state:

1. the structure of the grid
2. the previous path of the agent through the grid
3. the previous path of the other agent through the grid
4. information about final state

To reach one of the defined terminal conditions in training, the agent required 1363 episodes.

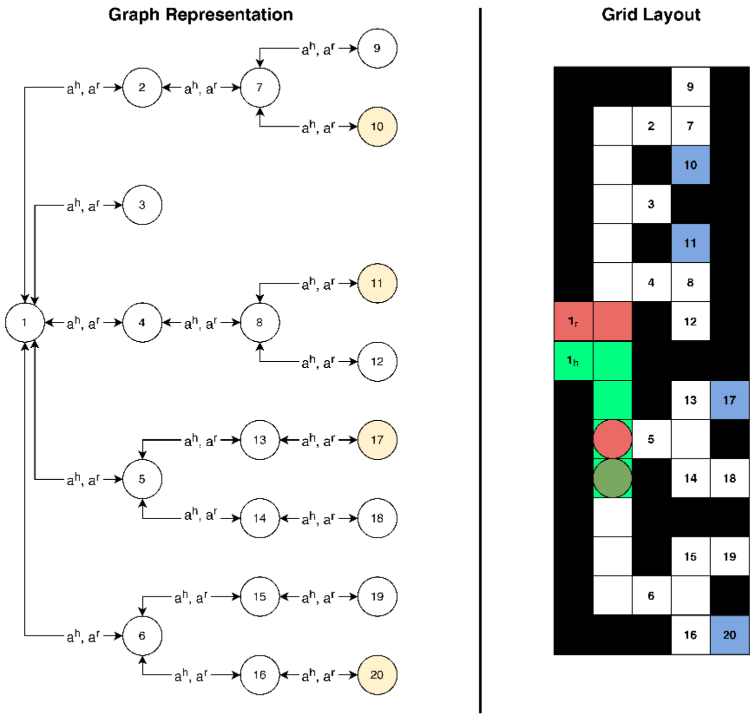


FIGURE 6. Grid Layout based on the graph representation used by Alessio et al. [25]]

Dos Santos et al. [26] consider task planning for mobile flying robots regarding the transportation and assembly of truss components. Since the authors consider not only the assembly planning but also the control of the robots, they divide their approach into two phases. First, the control of the robot is trained, followed by the efficient planning of the task steps. This includes the sequence of the individual control commands, the correct selection of components regarding the structural properties of the assembly, as well as the sequence of the assembly steps. For the analysis of the work, the first part of the approach is primarily considered. The sum of the duration for adjustments to the height, x-z position and orientation of the robot is used as the cost function. The more adjustments are necessary, the lower the reward for the agent. A lower and upper limit for the reward avoids outliers. The environment is represented by a uniform grid with a side length of 7 cm per element. As an action, the agent selects an entry from an action matrix. The entries in this matrix consist of position and rotation entries for each component. A construction algorithm checks the structural integrity of the construction to identify incorrect sequences. At the same time, a path planning algorithm based on the A* algorithm evaluates possible flight paths, as these adapt dynamically as the assembly is built. After training, the learned behaviour is successfully transferred to a real environment by building a 3-story tower, requiring approximately 900 episodes during training.

Watanabe et al. [27] consider the case of efficient assembly sequence generation for a two-armed robot. The correct assembly of rectangular building blocks serves as the use case. The authors assume, that an assembly step is always reversible. Therefore, they follow the assembly by disassembly approach. The efficiency of an assembly is determined by the total assembly time. Here, the duration of an assembly step is described by the duration of movement of the robot's hands. The description of the state also consists of a binary list, whereby the length of the list corresponds to the number of components and the value in the list describes the disassembly of a respective component. The action is described by a tuple with two integer values. Each value represents the disassembly of a component with the corresponding number. The 0 stands for the disassembly of no component. The reward is split into three parts:

1. the agent receives a reward if the disassembly is complete and based on the duration of the entire disassembly.
2. the agent receives a reward if the disassembly of a component was completed.
3. the agent receives a negative reward if the disassembly did not take place.

A collision check is carried out to check whether a disassembly is faulty. After training an assembly arrangement, transfer learning is used to train the model on further assembly arrangements.

In their work, Winter et al. [28] focus on interactive reinforcement learning (IRL) between humans and robots. The authors present an approach to train a RL Agent in a virtual environment in cooperation with an external human actor in a real environment (figure 7). The authors aim to accelerate the training performance. Communication from the human to the RL agent is based on spoken language. In the virtual environment, the robot's actions are evaluated in compliance with selected limitations. A Greedy Selector chooses the most suitable action, based on the evaluations received. The robot's real counterpart executes the action in the real environment, so the human can observe its behaviour and can verbally define new limitations for the robot to influence the selection of actions. The aim is to learn not only the sequence of components, but also the robot's behaviour during assembly. A binary list is used to represent the status of the components. The number of actions is 5 and these are defined by a discrete value. The length of the assembly should be minimized. Accordingly, a negative reward is awarded for each action. If a component is successfully assembled, a positive reward is awarded as well. The authors use the Cranfield benchmark problem [39], which consists of the assembly of a pendulum. After approximately 50 episodes of simulated training, the number of constrain violations is approximately 0. The execution of the real user study shows that the approach can respond to user feedback by adapting behaviour within a few episodes.

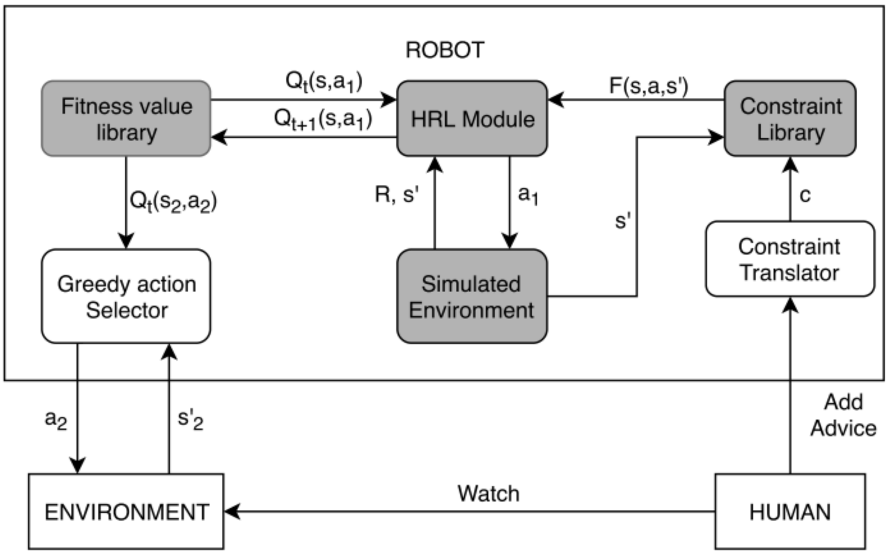


FIGURE 7. Environment setup used by Winter et al. [28]

In [29], Winter et al. focus on the optimization of a given assembly sequence. A virtual environment is used to train physically possible assembly sequences. Subsequently, the trained behaviour is transferred to a real environment to identify the fastest of the identified assembly sequences. The virtual environment is described as a digital counterpart. The authors compare different 3D engines in terms of their collision accuracy, performance, and flexibility. In the first phase, a negative reward is awarded for an error in the assembly sequence. Plausible assembly sequences have a corresponding reward of 0. The states of the components are saved in a binary list. For phase 2, the description of the status from phase 1 is adopted and expanded according to the last skill performed. The reward is assigned based on the execution time multiplied with -1. Carrying out the case study, the authors state that training the real robot and the digital twin in parallel speeds up the training. In fact, the parallel approach converges about 7 times faster and finds an optimum after about 100 episodes. The authors use the Cranfield benchmark problem [39] as well.

Cebulla et al. [30] aim to use the Monte Carlo tree search algorithm (MCTS) and Q-functions to solve the ASP problem, by leveraging the assembly by disassembly approach. The concept makes use of pre-training to create a model that is easily adaptable to specific use cases. The model is trained to minimize the number of directional changes needed to assemble the product and to maximize the accessibility of each component during the assembly process. A graph neural network is used as the model architecture. To model the collision-free space in the six axis directions for each component, a 2 ½ D matrix is used. The matrix is also used to check the stability of the components regarding gravity, by checking the collision-free space on the negative y-axis, and to determine the accessibility of the components. In addition to accessibility, the number of directional changes during assembly should also be reduced. To minimize the directional changes, the disassembly direction vector of each component is selected so that it is parallel to the direction vector from the previous step, if possible. The reward is the sum of the disassembly direction vectors of the selected component that are parallel to the direction vectors of the previous steps.

The approach is validated with two assemblies, with 38 and 58 components. The authors can show that the use of pre-trained Q-functions in combination with MCTS provides better results than the use of an unmodified MCTS, as well as the use of a Q-function specifically trained to the use case in combination with MCTS.

Guo et al. [31] use a connector-linked model, which is formulated as a graph, to model connector elements to be assembled. For training, the graph is formulated as a matrix for each connector. Each row contains the ID of the connector, the ID of the connector, which must be assembled beforehand, the connector type, the currently selected tool, and the current direction of the assembly. As the last two properties change during assembly, a new row with the corresponding values is added to the matrix for each plausible combinatorial possibility. Accordingly, the first three entries of each row are identical. Additionally, a binary value for each connector describes if the corresponding part can be selected. The List of matrices and binary values represent the state of the assembly.

The objective function aims to minimize the changes of the connector type, the assembly tool and assembly direction. The Reward is a weighted sum of these factors. The authors used 8 different connector and tool types in the study case. The total number of connectors is 20, 30, 40 and 50. The authors trained for 30.000 steps, which took 20-40h, depending on the number of connectors. After training, the model finds near optimal solutions in 0.6s in comparison to heuristic algorithms, which require 2-8 minutes.

B. Disassembly Sequence Planning

In [32], Bi et al. use Q-learning to search for a disassembly sequence of an assembly. To describe the assembly, the authors use an AND/OR graph that represents the relationships between the individual components in the assembly. A separate reward is defined for each component, which is awarded when the corresponding component is disassembled. To describe the status, the AND/OR graph is represented by a matrix M. Each row in M is a single state and has the following properties:

1. A start and a final state are defined to describe the part ranking.
2. An identifier of a disassembled component at the corresponding position.

The authors compare three use cases. A washing machine with 6 components and 15 states. A rotor assembly with 7 components and 17 states. And a hammer drill with 22 components and 63 states. The authors also compare Q-learning and GA in the use cases, whereby the RL approach not only converges significantly faster but also offers more room for improvement, according to the authors. The Q-learning approach requires about 8000 steps to find an optimal solution, for use case 3.

Chen et al. [33] also use Q-learning to find a disassembly sequence of selected smartphone components. The authors form the reward from the disassembly duration $R(t)$ and the disassembly of a component $R(m)$, whereby a corresponding reward is defined for each component. In addition to the actual execution time, the time required to change a tool is also used to determine the disassembly time. The resulting value t_i is finally offset against the value factor Z, which in the application case is 100.

$$R = \frac{Z}{t_i} + R(m)$$

If the agent performs an invalid disassembly, it is penalized with a negative reward. A list with n elements for all n components is used to describe the status. The same applies to the description of the action. In this way, the authors create a state-action-reward matrix. In the study case, the disassembly of a smartphone with 11 components and 5 tools is considered. The number of iterations performed is 50 until an optimum disassembly sequence was found. Yang et al. [34] consider the robot DSP problem, i.e. finding the optimal disassembly sequence using a robot. The reward is awarded according to the following equation:

$$r_i = T_{mx} - T_{tool\ switch} - T_{direction\ switch} - T_{moving}$$

T_{mx} represents the maximum duration for each disassembly step, T_{moving} is the time the robot needs to move between two disassembly points. $T_{tool\ switch}$ describes the duration of the tool switch and $T_{direction\ switch}$ is the duration of the robot to adjust its direction of movement. The state description consists of two lists D^k and C^k, where k corresponds to the number of components to be disassembled. D^k is a binary list for representing the components that have already been dismantled. The list C^k is one-hot encoded and describes the component currently being disassembled by the agent/robot. A discrete number from 1 to k is used as the action for the disassembly of the corresponding component. The lists are concatenated in a predefined order to describe the status. The case study is based on a double coupling shaft with 21 components. The authors compare the performance of the DRL approach with a bee algorithm and a genetic algorithm. For all three approaches, 300 iterations were performed in training. The DRL approach finds the fastest disassembly sequence, while the genetic algorithm determines the slowest of all three sequences.

Zhao et al. [35] argue that the commonly used data structures, such as an AND/OR graph or a disassembly tree, are not capable of mapping the constraints of the components to each other. The authors therefore present a new data structure, based on a hybrid graph model (HGM) data structure. The authors extend the data structure with three matrices to achieve a higher degree of detail regarding dynamic changes to the disassembly of individual instances:

1. The precedence matrix describes the precedence relationships among the components. A distinction is made among AND & OR relationships. The AND relationship describes that a part can only be dismantled once all previous dismantling parts have been removed. The OR relationship means that a part can be removed after one of its previous disassembly parts has been removed.
2. The contact matrix maps the contact relationships among the components. It includes both direct contact constraints and indirect contacts among parts.
3. The level matrix comprises the disassembly level of each component, which is determined by the leading edges of the components. Components that can be disassembled simultaneously are located on the same level.

If components are missing, an algorithm adjusts the data structure accordingly. A state is described using the three matrices and the previous action. An action is a discrete value that represents the disassembly of a component that is still present. The authors refer to this as a multi-level selective disassembly hybrid graph model (MSDHGM). The reward is determined by the required disassembly time and the profit of the disassembled component. In addition, a negative reward is awarded if the disassembly direction or the disassembly tool is changed. If components are disassembled that minimize the total number of disassembly steps due to their dependencies, this has a positive influence on the reward. To validate the approach, the authors use an assembly with 26 components. The duration of the training is not specified. The authors compare the performance of a DQN with that of a non-dominated sorting genetic algorithm and an artificial bee colony algorithm. The evaluation of the generated sequences shows that the DQN determines solutions that are closer to the optimal ones.

Cui et al. [36] also argue that common data structures for mapping disassembly sequences and similar dependencies of components are not sufficient to adequately cover the dynamic problems for EOL products. The authors suggest that the wear, deformation, corrosion, and possible fractures of the components should be assessed by experts before disassembly. The assessments are weighted and used to determine the disassembly time. The assessments are also used to check whether possible disassembly sequences can now be hindered. The description of the status is made up of the information about the disassembly status (binary list) of each element and the component currently to be disassembled. To calculate the reward, the basic disassembly duration of a component, the duration of the tool change, the duration of the agent's movements and the extra duration determined from the evaluations are added together. The resulting total is subtracted from a specified maximum value. To verify the effectiveness of proposed method, a double coupling shaft with 21 parts is used in a case study. Components 5 and 20 may have faulty properties which have been evaluated by an expert. In addition, the authors compared the DQN approach with other algorithms, such as an improved GA or a duelling DQN. The normal DQN finds the fastest disassembly sequence with the shortest training duration of 814 seconds. The time to calculate the sequence is less than one hundredth of a second after the training is complete.

Allagui et al. [37] present an approach for generating disassembly sequences using RL as well. The concept is based on two main steps. Firstly, the processing of a collision matrix. This is based on CAD data of the components and describes the possible collision of each component in six axis directions. In contrast to Guo et al. [31], the values of the collision matrix are binary coded and therefore do not represent the distance to other components. The second step is the execution of the RL algorithm.

To determine the fitness function, the authors consider the following properties:

1. the normalized volume of the selected component (smaller is better)
2. the normalized duration of a disassembly step (smaller is better)
3. the change in the disassembly direction (1: no change, 0.5: 90°, 0: 180°)
4. the change of the tool, binary encoded
5. whether the selected component is a wear part, also binary encoded

The factors are expressed as a weighted sum. To determine the reward, the value is multiplied by a factor of 100 if a component has been disassembled. Otherwise, the reward is -1, if it’s not physically possible to remove the selected part. As an experiment, the authors choose an assembly with five components. There are a total of 22 possible disassembly sequences.

In [38], Liu et al. aim to disassemble an EOL product with a robotic arm, under the assumption that some components of the product are missing. To determine the missing parts, several depth cameras are used and evaluated with a deep learning model. Afterwards, a deep Q-learning model is used to select the optimum action for disassembly, based on the condition of the product. In addition, a digital twin of the disassembly robot is connected to the system and linked to the real scene. The status of the assembly is represented as a vector consisting of three elements:

1. disassembled components (binary encoded)
2. the currently disassembled component (one hot encoded)
3. missing components (binary coded)

The reward is calculated using a constant value minus the required time to assemble the selected product. The authors used two assemblies with 17 and 23 components to carry out the experiment. The training includes approximately 11.000 and 14.000 episodes. Afterwards, a near optimal solution can be found in half a second for the first assembly and 1.5 seconds for the second assembly.

IV. DISCUSSION

In this section, the presented works are analysed regarding the defined research questions to give recommendations on how to apply reinforcement learning to the domain of (dis-)assembly sequence planning.

A. Reward Definition

The most frequently considered factor for the design of the reward function is the required duration of executing the proposed (dis)assembly sequence. This is usually done directly, measuring the required time [16,17,24,27,33–36], or indirectly taking into account the number of steps [22,25,28]. In the latter case, a negative reward is awarded for each planned step, while the determination of the time usually considers multiple factors, such as the time required to execute a step, the time required for any tool changes or, if available, the additional time required to align the robot between the individual steps. In addition, some authors recommend offsetting the accumulated time against a maximum possible time. This can be done either by normalization [16,17] or a simple subtraction [34].

Another property for awarding both negative and positive rewards is the state of the assembly. The achievement of (final) valid states is rewarded with a positive reward [21–23,25,29], while the achievement of invalid states is penalized with a negative reward [25,29,33]. The determination of the validity of a state is derived from previously defined conditions for the assembly. For problems in disassembly sequence planning, some authors award the disassembly of certain components [32,33,35]. The amount of the reward depends on the selection of the component.

More specific rewards, which relate to the respective configurations of the assembly, are also presented. For example, the number of required support structures is minimized [19], the collision-free space around the selected component is maximized [20] or the stability of the assembly is maximized [24]. Robotic use cases consider additional application-specific rewards. On the one hand, the power consumption of the robot is minimized [24]. In a second work, the number of adjustments in the motion sequence of an autonomous flying robot is minimized [26]. For use cases in the field of human-centric approaches, two studies suggest incorporating feedback from the technician into the agent's reward [17,18]. This enables the human to give feedback according to the ergonomics of the individual construction steps to optimize them accordingly.

Based on this analysis, we recommend considering three different factors when defining the reward function:

1. The duration required for executing the (dis)assembly, either determined via the accumulated time or via the number of required steps. The latter is independent of the number of elements in the assembly and can therefore be used for assemblies of any size without additional adjustments. It is advisable to consider as many factors as possible, namely the time required executing a step, the time for any tool changes and, if applicable, the additional time required for the alignment of the robot between the individual steps. If these times are known, it can be recommended to choose this approach, due to stating the execution time is more precise. To increase the robustness during training, the time required should be normalized with a maximum duration.
2. Awarding achieving desired assembly states. On the one hand, desired configuration can be rewarded, or invalid states can be penalized. In this way, the conditions imposed on the assembly can also be mapped in the reward. In the context of disassembly, this also includes the awarding of a specific reward for the disassembly of certain components. Alternatively, desired properties

of the assembly, such as free space, stability, or human feedback, can be used. These can be determined, for example, by a simulation environment or in the field, while the determination of the reward for desired conditions must be defined in advance.

3. A reward adapted to the robot, such as the required energy or number of actions.

The final reward can be defined as the weighted sum of the three parts. The weighting of the individual parts should be adapted to the use case.

B. State Definition

The representation of the state is essential for the agent to understand its environment and be able to select suitable actions. Based on the action, the identified works can be divided into two approaches affecting the state representation. The agent can be trained to select the next part to be (dis)assembled on a set of possible parts, or it can be trained to decide if a given part should be (dis)assembled. If the latter approach is chosen, the state representation must cover the current state of the assembly, as well as the selected part, which is achieved by a discrete identifier for each part [36] or a one-hot encoded list [34].

The simplest representation of the assembly state is the subdivision into possible configuration states. The state description can thus be expressed as a discrete number [22,23]. A more comprehensive and proven description of the state of the assembly is a binary list. This list contains an entry for each component of the module, which describes the status of the component. This entry can either have the value 0 or 1. This simplistic representation is used in most publications [16,17,33,34,36]. However, it also leaves out some information. Therefore, some authors extend the list with more information like the currently selected tool expressed as a discrete number [16,17].

Another proven data structure is a connectivity matrix. It contains a row and column for each component of the assembly. The value at the position (i,j) describes the connection between components i and j. This representation is particularly useful for components with several possible connections, such as truss systems [19,24]. Spanning a grid to map the surroundings is also presented. Here, the space around the module is divided into a grid. Each grid field can assume a value that describes the state of the field [26]. Alessio et al. use the flexibility of this approach to represent configuration states as grid fields [25]. Valid, invalid, and final configurations are represented with different values.

Especially in disassembly sequence planning, the representation of the assembly is often represented as an AND/OR graph [32,35]. Accordingly, some authors deal with the processing of the data structure into a suitable state description of the assembly for the agent. Zhao et al. propose their MSDHGM data structure for this purpose [35]. The use of camera images as a status description of the assembly and the component to be assembled is also considered [21]. This approach is particularly flexible in its application, but also time-consuming for training the agent. The authors therefore suggest initializing the network with a pre-trained network and using transfer learning. In general, it can be said that when working in the field of robotics, the state description is divided into two parts. On the one hand, the state of the assembly is described, and on the other, the state of the robot [24,34].

The analysis of the state descriptions shows that the use of a binary list describing the state description of an assembly is sufficient in many cases. Due its simplistic one-dimensionality, the training time is limited because of the fewer combinatorial possibilities to be considered. However, this representation should not be used if components can have several possible connections, e.g. in truss systems. In this case, the use of a connectivity matrix is recommended instead. It is also possible to represent the state of the assembly as a grid. This approach is particularly flexible to use. However, when mapping a large space around the assembly, the size of the grid also tends to become very large, which therefore increases the number of combinatorial possibilities and corresponds in an increased training time.

Although the use of an image to describe the status of the assembly and, if applicable, the component to be installed was also used, this approach is proving to be time-consuming for training the agent. Additionally, it must be ensured that all aspects relevant to the agent's decision are covered in the images, so it may be necessary to take several images of the assembly from different perspectives. If this approach is chosen, it is advisable to train the network with an image data set in advance to shorten the training.

The descriptions mentioned so far refer exclusively to the state of the parts or assembly. Depending on the use case, it may also be necessary to consider the state of a robot. General statements are difficult to define. However, factors that are included in the determination of the reward must also be taken into account in the state description of the agent.

C. Potential Research Opportunities

Some studies compare the performance of RL with genetic algorithms in the respective application [17,31,32,34–36]. RL methods are often faster in determining the sequences as well as in the performance of the generated sequences regarding the defined costs. However, it should be noted that, depending on the complexity of the use case, training is very time-consuming. Therefore, many of the presented works limit the number of combinatorial possibilities in the studied use cases [22,23,25].

However, the application of concepts such as curriculum learning, parameter transfer or fine-tuning for specific use cases seems promising, but are not yet very widespread [19,35]. For this reason, we argue that research into highly general models through training on many different use cases with subsequent fine-tuning on the respective use case has great potential. Central to the development of such a model will be concepts such as parameter transfer and curriculum learning. Accordingly, more research is needed into the application of such a model in a variety of different industrial application areas.

V. CONCLUSION

In conclusion, the analysis of reinforcement learning applications in (dis-)assembly sequence planning provides valuable insights for the effective implementation of this technology. The recommended approach for defining reward functions involves considering three key factors: the duration of (dis)assembly execution, the achievement of desired assembly states, and rewards adapted to specific robotic parameters. By incorporating these factors into a weighted sum, tailored to the use case, a robust reward function can be devised. The examination of state representations highlights the suitability of a binary list for many cases due to its simplicity and efficient training time. However, the use of a connectivity matrix is advised for scenarios involving components with multiple possible connections, such as truss systems. Alternatively, a grid-based representation offers flexibility but requires careful consideration of the combinatorial possibilities to avoid increased training times.

While the use of images to describe assembly status is flexible, it proves time-consuming and necessitates comprehensive coverage of decision-relevant aspects. Overall, the state description must not only encompass the assembly but also consider factors relevant to the robot's state, aligning with the reward determination. This comprehensive understanding of reward functions and state representations provides a solid foundation for the successful integration of reinforcement learning in (dis-)assembly sequence planning across diverse applications and scenarios. The limited number of identified publications indicate, that the utilization of reinforcement learning methods for the generation of assembly and disassembly sequences is an underexplored subject. Therefore, possible areas of interest for future works were outlined, to guide future research in the described field.

ACKNOWLEDGMENT

This research received no external funding. The authors declare no conflict of interest.

REFERENCES

[1] Boothroyd, G.; Alting, L. Design for Assembly and Disassembly. CIRP Annals 1992, 41, 625–636, doi:10.1016/S0007-8506(07)63249-1.

[2] Fazio, T.L. de; Edsall, A.C.; Gustavson, R.E.; Hernandez, J.; Hutchins, P.M.; Leung, H.-W.; Luby, S.C.; Metzinger, R.W.; Nevins, J.L.; Tung, K.; et al. A Prototype of Feature-Based Design for Assembly. Journal of Mechanical Design 1993, 115, 723–734, doi:10.1115/1.2919261.

[3] Wang, K.; Li, X.; Gao, L. Modeling and optimization of multi-objective partial disassembly line balancing problem considering hazard and profit. Journal of Cleaner Production 2019, 211, 115–133, doi:10.1016/j.jclepro.2018.11.114.

[4] Homem de Mello, L.S.; Sanderson, A.C. A correct and complete algorithm for the generation of mechanical assembly sequences. In Proceedings / 1989 IEEE International Conference on Robotics and Automation, [May 14 - 19, 1989, Scottsdale, Arizona]. Proceedings, 1989 International Conference on Robotics and Automation, Scottsdale, AZ, USA, 14-19 May 1989; IEEE Computer Society Press: Washington, DC, 1989; pp 56–61, ISBN 0-8186-1938-4.

[5] Jiménez, P.; Torras, C. An efficient algorithm for searching implicit AND/OR graphs with cycles. Artificial Intelligence 2000, 124, 1–30, doi:10.1016/S0004-3702(00)00063-1.

[6] Wang, D.; Shao, X.; Ge, X.; Liu, S. A hybrid technology for assembly sequence planning of reflector panels. AA 2017, 37, 442–451, doi:10.1108/AA-12-2016-171.

[7] Tseng, H.-E.; Chang, C.-C.; Lee, S.-C.; Huang, Y.-M. A Block-based genetic algorithm for disassembly sequence planning. Expert Systems with Applications 2018, 96, 492–505, doi:10.1016/j.eswa.2017.11.004.

[8] Hong, D.S.; Cho, H.S. A neural-network-based computational scheme for generating optimized robotic assembly sequences. Engineering Applications of Artificial Intelligence 1995, 8, 129–145, doi:10.1016/0952-1976(94)00068-X.

[9] Sutton, R.S.; Barto, A. Reinforcement learning: An introduction, Second edition; The MIT Press: Cambridge, Massachusetts, London, England, 2018, ISBN 9780262352703.

[10] Arulkumaran, K.; Deisenroth, M.P.; Brundage, M.; Bharath, A.A. Deep Reinforcement Learning: A Brief Survey. IEEE Signal Process. Mag. 2017, 34, 26–38, doi:10.1109/msp.2017.2743240.

[11] Goodfellow, I.; Courville, A.; Bengio, Y. Deep learning; The MIT Press: Cambridge, Massachusetts, 2016, ISBN 9780262337373.

[12] D. Silver. Lectures on reinforcement learning, 2015.

[13] Feurer, M.; Hutter, F. Hyperparameter Optimization. In Automated machine learning: Methods, systems, challenges; Hutter, F., Kotthoff, L., Vanschoren, J., Eds.; Springer: Cham, Switzerland, 2019; pp 3–33, ISBN 9783030053178.

[14] A. D. Laud. Theory and application of reward shaping in reinforcement learning, 2004.

[15] Aria, M.; Cuccurullo, C. bibliometrix : An R-tool for comprehensive science mapping analysis. Journal of Informetrics 2017, 11, 959–975, doi:10.1016/j.joi.2017.08.007.

[16] Neves, M.; Vieira, M.; Neto, P. A study on a Q-Learning algorithm application to a manufacturing assembly problem. Journal of Manufacturing Systems 2021, 59, 426–440, doi:10.1016/j.jmsy.2021.02.014.

[17] Neves, M.; Neto, P. Deep reinforcement learning applied to an assembly sequence planning problem with user preferences. Int J Adv Manuf Technol 2022, 122, 4235–4245, doi:10.1007/s00170-022-09877-8.

[18] Giorgio, A. de; Maffei, A.; Onori, M.; Wang, L. Towards online reinforced learning of assembly sequence planning with interactive guidance systems for industry 4.0 adaptive manufacturing. Journal of Manufacturing Systems 2021, 60, 22–34, doi:10.1016/j.jmsy.2021.05.001.

[19] Hayashi, K.; Ohsaki, M.; Kotera, M. Assembly Sequence Optimization of Spatial Trusses Using Graph Embedding and Reinforcement Learning. Journal of the International Association for Shell and Spatial Structures 2022, 63, 232–240, doi:10.20898/j.iass.2022.016.

[20] Kitz, K.; Thomas, U. Neural dynamic assembly sequence planning. In 2021 IEEE 17th International Conference on Automation Science and Engineering (CASE). 2021 IEEE 17th International Conference on Automation Science and Engineering (CASE), Lyon, France, 23–27 Aug. 2021; IEEE, 2021; pp 2063–2068, ISBN 978-1-6654-1873-7.

[21] . Zhao, M.; Guo, X.; Zhang, X.; Fang, Y.; Ou, Y. ASPW-DRL: assembly sequence planning for workpieces via a deep reinforcement learning approach. AA 2019, 40, 65–75, doi:10.1108/AA-11-2018-0211.

[22] Antonelli, D.; Zeng, Q.; Aliev, K.; Liu, X. Robust assembly sequence generation in a Human-Robot Collaborative workcell by reinforcement learning. FME Transactions 2021, 49, 851–858, doi:10.5937/FME2104851A.

[23] Antonelli, D.; Aliev, K. Robust assembly task assignment in Human Robot Collaboration as a Markov Decision Process problem. Procedia CIRP 2022, 112, 174–179, doi:10.1016/j.procir.2022.09.068.

[24] Yin, J.; Chen, M.; Zhang, T. Optimization Algorithm for Cooperative Assembly Sequence of Truss Structure Based on Reinforcement Learning. In Intelligent Robotics and Applications; Liu, X.-J., Nie, Z., Yu, J., Xie, F., Song, R., Eds.; Springer International Publishing: Cham, 2021; pp 474–484, ISBN 978-3-030-89091-9.

[25] Alessio, A.; Aliev, K.; Antonelli, D. Robust Adversarial Reinforcement Learning for Optimal Assembly Sequence Definition in a Cobot Workcell. In Advances in Manufacturing III; Trojanowska, J., Kujawińska, A., Machado, J., Pavlenko, I., Eds.; Springer International Publishing: Cham, 2022; pp 25–34, ISBN 978-3-030-99309-2.

[26] dos Santos, S.R.B.; Givigi, S.N.; Nascimento, C.L. Autonomous construction of structures in a dynamic environment using Reinforcement Learning. In 2013 IEEE International Systems Conference (SysCon). 2013 7th Annual IEEE Systems Conference (SysCon), Orlando, FL, 15–18 Apr. 2013; IEEE, 2013; pp 452–459, ISBN 978-1-4673-3107-4.

[27] Watanabe, K.; Inada, S. Search algorithm of the assembly sequence of products by using past learning results. International Journal of Production Economics 2020, 226, 107615, doi:10.1016/j.ijpe.2020.107615.

[28] Winter, J. de; Beir, A. de; El Makrini, I.; van de Perre, G.; Nowé, A.; Vanderborght, B. Accelerating Interactive Reinforcement Learning by Human Advice for an Assembly Task by a Cobot. Robotics 2019, 8, 104, doi:10.3390/ROBOTICS8040104.

[29] Winter, J. de; EI Makrini, I.; van de Perre, G.; Nowé, A.; Verstraten, T.; Vanderborght, B. Autonomous assembly planning of demonstrated skills with reinforcement learning in simulation. Auton Robot 2021, 45, 1097–1110, doi:10.1007/s10514-021-10020-x.

[30] Cebulla, A.; Asfour, T.; Kröger, T. Efficient Multi-Objective Assembly Sequence Planning via Knowledge Transfer between Similar Assemblies. In 2023 IEEE 19th International Conference on Automation Science and Engineering (CASE). 2023 IEEE 19th International Conference on Automation Science and Engineering (CASE), Auckland, New Zealand, 26–30 Aug. 2023; IEEE, 2023; pp 1–7, ISBN 979-8-3503-2069-5.

[31] Guo, K.; Liu, R.; Duan, G.; Liu, J.; Cao, P. Research on dynamic decision-making for product assembly sequence based on Connector-Linked Model and deep reinforcement learning. Journal of Manufacturing Systems 2023, 71, 451–473, doi:10.1016/j.jmsy.2023.09.015.

[32] Bi, Z.; Guo, X.; Wang, J.; Qin, S.; Qi, L.; Zhao, J. A Q-Learning-based Selective Disassembly Sequence Planning Method. In 2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC). 2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Prague, Czech Republic, 09–12 Oct. 2022; IEEE, 2022; pp 3216–3221, ISBN 978-1-6654-5258-8.

[33] Chen, Z.; Li, L.; Zhao, F.; Sutherland, J.W.; Yin, F. Disassembly sequence planning for target parts of end-of-life smartphones using Q-learning algorithm. Procedia CIRP 2023, 116, 684–689, doi:10.1016/j.procir.2023.02.115.

[34] Yang, C.; Xu, W.; Liu, J.; Yao, B.; Hu, Y. Robotic Disassembly Sequence Planning Considering Robotic Movement State Based on Deep Reinforcement Learning. In 2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design (CSCWD). 2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Hangzhou, China, 04–06 May 2022; IEEE, 2022; pp 183–189, ISBN 978-1-6654-0527-0.

[35] Zhao, X.; Li, C.; Tang, Y.; Cui, J. Reinforcement Learning-Based Selective Disassembly Sequence Planning for the End-of-Life Products With Structure Uncertainty. IEEE Robot. Autom. Lett. 2021, 6, 7807–7814, doi:10.1109/LRA.2021.3098248.

[36] Cui, J.; Yang, C.; Zhang, J.; Tian, S.; Liu, J.; Xu, W. Robotic disassembly sequence planning considering parts failure features. IET Collab Intel Manufact 2023, 5, doi:10.1049/cim2.12074.

[37] Allagui, A.; Belhadj, I.; Plateaux, R.; Hammadi, M.; Penas, O.; Aifaoui, N. Reinforcement learning for disassembly sequence planning optimization. Computers in Industry 2023, 151, 103992, doi:10.1016/j.compind.2023.103992.

[38] Liu, J.; Xu, Z.; Xiong, H.; Lin, Q.; Xu, W.; Zhou, Z. Digital Twin-Driven Robotic Disassembly Sequence Dynamic Planning Under Uncertain Missing Condition. IEEE Trans. Ind. Inf. 2023, 19, 11846–11855, doi:10.1109/TII.2023.3253187.

[39] Collins, K.; Palmer, A.J.; Rathmill, K. The Development of a European Benchmark for the Comparison of Assembly Robot Programming Systems. In Robot Technology and Applications: Proceedings of the 1st Robotics Europe Conference Brussels, June 27-28, 1984; Rathmill, K., MacConaill, P., O'Leary, P., Browne, J., Eds.; Springer: Berlin, Heidelberg, 1985; pp 187–199, ISBN 978-3-662-02442-3.