THE INFLUENCE OF REPRESENTATION ON SYSTEM INTERPRETATION: A SEARCH FOR MOST COMMON SET PARTITIONS

Murphy, Alexander R. (1);
Patel, Apurva R. (1);
Zorn, Stefan (2);
Gericke, Kilian (2);
Summers, Joshua D. (1)

1: University of Texas at Dallas;
2: University of Rostock

ABSTRACT

During engineering design, different representations are used to convey information about a systems' components, functionality, spatial layout, and interdependencies. These varying representations may have an impact on the interpretation of a system and consequently the decision-making process. This paper presents a research study that tries to capture these different interpretations by investigating how designers divide a system into subsystem clusters. These subsystem clusters can be considered partitions of a set-in combinatorial mathematics. Given designers' subsystem clusters for three products across three representation modalities, three different analysis methods for finding the most likely partition from observed data are presented. Analysis shows that the Variation of Information analysis method gives the most coherent and consistent results for the search of a most likely cluster. In addition, differences in clustering behaviour are observed based on representation modality. These results show that the way an engineer or designer chooses to represent a system impacts how that system is interpreted, which has implications for the decision-making process during engineering design.

Keywords: Design theory, Design cognition, Design representation, Systems thinking, Human behaviour in design

Contact:
Murphy, Alexander R.
University of Texas at Dallas
United States of America
alexander.murphy@utdallas.edu

Cite this article: Murphy, A. R., Patel, A. R., Zorn, S., Gericke, K., Summers, J. D. (2023) 'The Influence of Representation on System Interpretation: A Search for Most Common Set Partitions', in *Proceedings of the International Conference on Engineering Design (ICED23),* Bordeaux, France, 24-28 July 2023. DOI:10.1017/pds.2023.343

1 INTRODUCTION

Systems thinking is critical to the engineering design process. With increasingly complex technology in a highly connected society, designers must be able to understand, interpret, and leverage complex socio-technical systems. Large scale problems such as resource scarcity and climate change demand engineering solutions that are innovative and sustainable. These challenges have prompted research on complex engineered systems and reemphasized the importance of systems thinking as a topic of research. Recent work has proposed system models for agricultural resource conservation (Barlow *et al.*, 2021), extended system design tools to consider users and context of the design problems (Liem, 2017), and proposed a multi-level system life cycle approach for the development of smart products and services (Forte, Göbel and Dickopf, 2021). Moreover, the rapid integration of physical and digital systems in manufacturing (Wichmann, Eisenbart and Gericke, 2019), and the emergence of cyber-physical-social systems (Zeng *et al.*, 2020) compels designers to manage multidisciplinary information at a system level. In addition to increasing design complexity, this also raises new questions for designing manufacturing systems (Waschull *et al.*, 2020).

Systems thinking is influenced by the manner used to represent an engineering system. Various methods for representing models of systems have been proposed and employed throughout literature. The Systems Modelling Language (SysML) has been proposed as a general modelling language that supports hardware and software elements (Friedenthal, Alan and Steiner, 2008). Early-stage representations focusing on functional decomposition have been a recurring topic of research (Hirtz *et al.*, 2002; Erden *et al.*, 2008; Summers, Eckert and Goel, 2017). System representations developed using set theory and graph-based approaches have also been used (Shai and Preiss, 1999; Buede and Miller, 2016). Given the many possible ways that a designer or engineer can choose to represent a system, it is important to understand how each representation modality impacts system understanding, influences communication, and interacts with existing mental models of the system. Despite these various representation tools, it is not well understood how modality impacts behaviour during the engineering design process.

This paper builds from the prior work and presents a research study that takes strides towards understanding the relationship between representation modality and modeling behaviour through an experimental study. This study leverages a subsystem clustering task with an associated design prompt to explore these relationships. Analysis and results focus on an exploration of methods to determine the most likely clustering pattern given participants' responses to the design task. The most likely clustering pattern refers to a partition that best fits the observed data set. These explored methods provide insight into the impact of representation modality on systems thinking behaviour across different systems. The primary research questions addressed in this work are:

RQ1: To what extent does representation modality affect system interpretation given a design task?

RQ2: What methods are most effective for determining a most likely partition of a set of elements that represent a system?

To address these research questions, an experimental study was designed to focus on clustering behaviour at the subsystem level. Three potential analysis methods for finding a most likely partition of the given systems are explored.

2 BACKGROUND

Functional decomposition in engineering design is described followed by an overview of mathematical concepts foundational to the analysis presented in this paper. The research presented in this paper builds from prior work (Murphy *et al.*, 2022; Patel *et al.*, 2022).

2.1 Functional decomposition

Functional decomposition is a common strategy in engineering design to represent a system or a product in terms of function instead of form (Otto and Wood, 2003; Pahl *et al.*, 2007; Dieter and Schmidt, 2009). A kind of functional decomposition called functional modeling is taught in engineering design classrooms to systematically represent engineering systems. Functional modeling research has focused on how to best teach students the technique (Nagel *et al.*, 2015), formalization of language used to describe functionality (Stone and Wood, 1999; Hirtz *et al.*, 2002), and its use for

https://doi.org/10.1017/pds.2023.343 Published online by Cambridge University Press

analogy-based design (Qian and Gero, 1996). The authors of this paper have also contributed to this body of work by applying functional modeling to research studies on mental models (Murphy *et al.*, 2019) and how individual differences impact modeling behaviour (Patel and Summers, 2021).
Functional models serve as one of the primary modalities for system representation implemented in the research study presented in this paper. The functional models in this work follow a traditional EMS (energy-material-signal) flow convention with functions described as verb-noun pairs. Abbott and Lough report on a study that connects physical system components to functional modeling through an educational tool (Abbott and Lough, 2007). The connection between physical components and functional decomposition is also present in this study as the chosen representation modalities.

2.2 Partitions of a set

Partitions of a set refer to a collection of subsets where each element appears within those subsets exactly once (Halmos, 1998). The subsets that comprise a partition of a set are often referred to as *blocks*, *parts*, or *cells* of the partition. In this paper, we will refer these subsets as *clusters*. Given a set of 5 elements, **Figure 1** shows all the possible partitions of the set that can be generated. Note that clusters with only a single element are without shading.

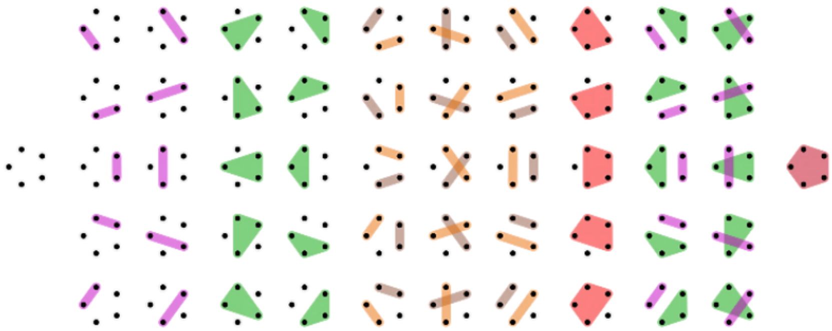

Figure 1: This shows the 52 possible partitions of a set with 5 elements (Piesk, 2011).

Participants in this study were similarly tasked with creating a set partition by indicating which elements from a system representation should be clustered together. In the design task, participants were instructed that an element could not belong to multiple clusters and each element had to be within a cluster even if the cluster only had a single element. The methodology section provides details on the design task and the study context.

2.2.1 Bell number

The Bell Number counts all possible unique partitions of a set with exactly n-elements (Kai, 1997). The Bell Numbers, typically denoted as B_n, form a sequence of integers (i.e. 1, 1, 2, 5, 15, 52, 203, 877, …). There are many mathematical applications for Bell Numbers such as the total rhyme schemes given n-lines of poetry or the distinct ways numbers can be factored into distinct primes (Gardner, 1978). A common procedural method for calculating the Bell Numbers is through the creation of a Bell Triangle (Gardner, 1978). Alternatively, a generating function $B(x)$ can be used that encodes the Bell Numbers as an infinite sequence of coefficients of a power series as shown in equation (1).

$$B(x) = \sum_{n=0}^{\infty} \frac{B_n}{n!} x^n = e^{e^x - 1} \quad (1)$$

For the representations presented in this study, the hair dryer, mixer, and toilet representations have Bell Numbers of 21147, 678570, and 4140 respectively. These Bell Numbers correspond to the total possible unique ways that participants could have completed the design task given the number of system elements in the representation. This list of all possible unique partitions will be generally referred to as a *Bell List* throughout this paper.

2.3 Variation of information

Several methods are available for comparing two partitions of a set. For instance, partitions may be cast as ordered pairs, which enables the use of pairwise distance measures such as cosine distance, Minkowski distance, or Hamming distance. Alternatively, partitions can be represented as text strings

https://doi.org/10.1017/pds.2023.343 Published online by Cambridge University Press

which can be compared using Levenshtein distance (Wagner and Fischer, 1974). Similarly, transforming the partitions into bipartite graphs can enable the use of graph theory-based approaches for comparing two partitions (Sanfeliu and Fu, 1983; Stauffer *et al.*, 2017). Set partitions discussed in this paper correspond to how participants understand a system with respect to their identification of subsystems. As such, the number of clusters and inclusion of elements in clusters are key aspects of interpretation. Using ordinal distance measures or comparing edit distance between strings may not be appropriate. While transformation to bipartite graphs preserves the clustering, it requires an additional layer of manipulation.

Alternatively, an information theory-based approach, using *Variation of Information*, has been proposed to compare partitions or clusters of sets (Meilă, 2007; Rossi, 2011). Variation of Information, or *shared information distance*, describes a method to determine the theoretical distance between two different partitions. This has been used to compare subspace clustering (Patrikainen and Meila, 2006) and to compare clustering methods for grouping design behaviours (Rahman *et al.*, 2018). This mitigates the need to transform set partitions into other representations and allows for direct comparison of information content in the partitions.

3 METHODOLOGY

For additional details and discussion about the study design, please refer to prior work on this topic (Murphy *et al.*, 2022; Patel *et al.*, 2022). The specific methods explored to find a most likely partition that represents the observed data set are described in Section 4 for organizational clarity.

3.1 Study context

This study was conducted at a university in northern Germany that offers a broad variety of undergraduate and graduate degree programs. The university emphasizes interdisciplinary collaboration throughout its programs. Participants for this study were students enrolled in an introductory engineering design course that lasts about 14 weeks. This course focuses on engineering skills such as free-hand sketching, engineering graphics, and descriptive geometry. The students come from degree programs such as mechanical engineering, industrial engineering, biomedical engineering, mechatronics, and some teaching degrees. Data collection occurred during the first week of the 14-week fall 2022 semester.

Of the students enrolled in this course, 72 students voluntarily participated in the research study. Of these 72 students, 61 identified as male and 11 identified as female. 68 of the 72 students reported that German is their native language. The study materials for data collection were translated from English into German by members of the research team fluent in both languages. 41 reported as being 1st year students, 14 as 2nd year students, 6 as 3rd year students, and 10 as being 4th year or more. Participants in the study beyond their 2nd year are likely transfer students or those who have changed degree programs, but this data was not collected.

3.2 Study design & procedure

In this study, three systems were selected for data collection. A hair dryer, a food mixer, and a standard toilet were chosen for a few reasons. First, these three systems are common household products such that most participants likely have some experience with them. Basic familiarity ensured that participants would have at least a rough mental model of the product and its functionality. Second, the three products relate to prior work. Specifically, the hair dryer has been used to measure mental model completeness as it relates to functional decomposition (Murphy *et al.*, 2019, 2020). The toilet was also used in this prior work, but as an example for how to complete the mental model task itself. The toilet was chosen because it shared little functional similarity to the other chosen products of that work (a clothes dryer and a vacuum cleaner) (Murphy, 2021). The research team has extensive experience studying the engineering design process through human subject studies, which informed the selection of the three products (hair dryer, mixer, toilet) presented in this paper.

Each of the three products was given to participants through three different representation modalities. Namely, a component graph, a function graph, and a function structure. The component graph shows an arrangement of internal components as images in locations where you might expect to see them if you were looking at a section-view of the product. The function graph replaces these components with functions (verb-noun pairs) in the same spatial locations as the component graph. These first two representations share spatial information but differ in the vocabulary used to describe elements. The

https://doi.org/10.1017/pds.2023.343 Published online by Cambridge University Press

third representation is referred to as a function structure. This representation preserves the functions from the function graph but rearranges them to form a more typical spatial layout expected from conventional functional modelling practices. In this study, the function structures were intentionally biased to have a left-to-right flow (with inputs at the left and outputs at the right) for consistency. Across all three representations, flow information (represented in the standard energy-material-signal notation) was preserved.

All three products were shown in the three representations resulting in a total of nine possible configurations (product-representation pairs). Each participant in the study received a packet with instructions for how to complete the assignment, examples for how to complete the clustering tasks, and three of the nine configurations. In other words, participants saw each of the products and each of the representations once. The three constructed packets were distributed to the participants randomly. For a complete description of the design task and prompt, see (Murphy *et al.*, 2022; Patel *et al.*, 2022).This paper reports on results from this research study with a total sample size of n = 72. Since the participants were divided into three groups randomly, sample sizes for each product-representation pair range from 21 to 27. The range of product-representation pairs sample sizes comes from participant packets that had to be removed from the data set because of errors in task completion or incomplete data.

4 ANALYSIS & RESULTS

Analysis of this data focuses on a search for a most likely partition for each product-representation pair to investigate if representation modality impacts systems thinking. This section discusses the three analysis methods used to determine a most likely partition. Each analysis method is described followed by the results of the analysis method with some interpretation. Limitations of each method are described. In-depth interpretation of these results is reserved for the discussion section.

To support analysis of partitions and clusters, a Cluster List and a Bell List was generated for each of the three products. As previously mentioned, the Bell List contains all possible partitions for a given set. Since the number of elements varied between products (between 8 and 11 elements), separate Bell Lists were generated. Similarly, a list of all possible subsets (clusters) was generated for each product. These Cluster Lists can be generated by recursive applications of n-choose-k over the list of elements. The number of all possible clusters for any given set follows $(2^n - 1)$, where n is the number of elements in the set. The number of elements, number of possible clusters, and the Bell Number for each product are presented in **Table 1**.

Table 1: Information about each product for data analysis

	Hair Dryer	Mixer	Toilet
# of Elements	9	11	8
Possible Clusters	511	2047	255
Bell Number	21147	678570	4140

For each product-representation pair, the Cluster List and Bell List were populated with the frequencies of each observed cluster and observed partition, respectively. A Bell Frequency Array and Cluster Frequency Array were then generated for each representation across all three products for a total of nine arrays for each.

4.1 Element cluster frequency

The first approach used to determine the most likely partition is based on the frequency of observed clusters. For each element in the set, the Cluster Frequency Array is filtered to include only the clusters that contain the given element. The filtered array is then queried for the maximum value, which is kept for creating the most likely partition. The process is then repeated for all elements in the set. Clusters with maximum frequency for each element were then used to generate a partition of the set. This process was repeated for each product-representation pair. In **Figure 2 (left)** below, the result of the analysis for the hair dryer component graph using the element cluster frequency method is shown.

https://doi.org/10.1017/pds.2023.343 Published online by Cambridge University Press

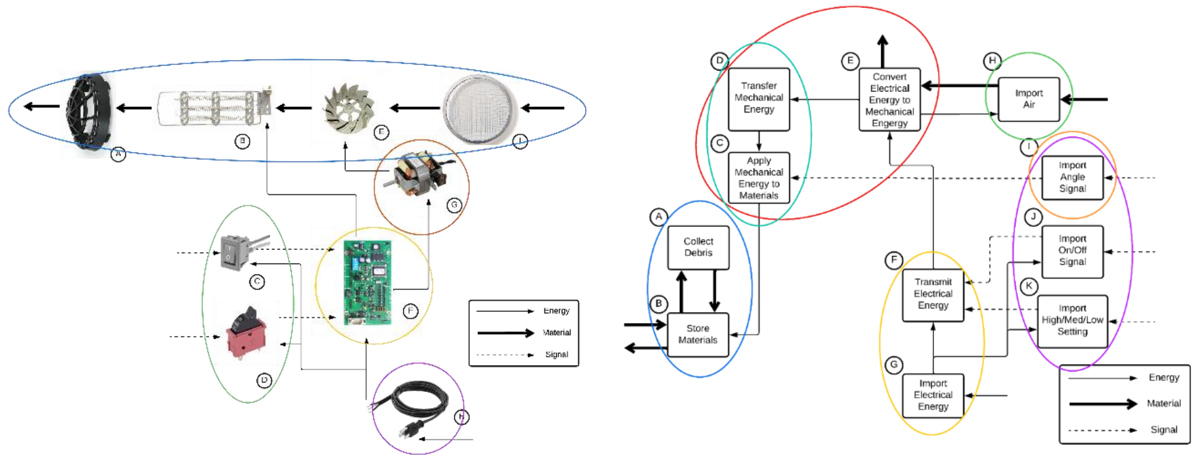

Figure 2. Result for the hair dryer component graph (left) and the mixer function graph (right) from the element cluster frequency analysis.

Initially, this method appears to give a coherent result. For the hair dryer component graph show in **Figure 2 (left)**, each element of this partition is within exactly one cluster as expected and the clusters include elements that are spatially adjacent to each other (see (Patel et al., 2022)). In other cases, partitions generated using this method contained elements that belonged to multiple clusters. The resulting partition for the mixer function graph is shown in **Figure 2 (right)**. This analysis method is unable to separate the clusters in a manner where they are distinct and non-overlapping. This occurs because an individual element is observed in multiple clusters where those clusters' frequencies are highest (i.e. C is most commonly with D {CD}, D is most commonly with C {CD}, and E is most commonly with D and E {CDE}). Inconsistencies can also occur when an individual element is in multiple clusters with equal frequency (i.e. A is found with B {AB} and C {AC} with equal frequency). While this analysis method does represent the observed data when considering each element's cluster frequency individually, it does not yield a most likely model that matches the design task itself (non-overlapping clusters). These inconsistencies were observed in 7 of 9 possible representation modalities across the three systems. Given this outcome, other analysis methods were explored.

4.2 Observed partition probability

The second approach evaluated the likelihood of observing any given partition in the Bell List given the observed clusters in participants' data. As shown in equation (2), a joint probability of all the clusters was computed to determine the partition probability.

$$P(X) = \prod_{i=1}^{N} P(x_i) \tag{2}$$

In equation (2), X represents a partition with clusters $x_1, x_2 \ldots x_N$, and the probability of each cluster is obtained from the Cluster Frequency Array. This is similar to the first approach where the frequency of clusters observed is used to construct a partition. However, in this case, the focus is on the whole partitions rather than the individual elements. This approach yields a probability for each possible partition rather than focusing on the observed cluster frequencies. The partition with the maximum probability is selected as the most likely partition for the set. **Table 2** shows the most likely partitions obtained for each product-representation pair.

Table 2: Most likely observed partitioning by probability

	Hair Dryer	Mixer	Toilet
Component Graph	{abegi}{cdfh}	{abc}{dehi}{fgjk}	{ad}{bcfg}{eh}
Function Graph	{abei}{cdfh}{g}	{ab}{cdefg}{hijk}	{abcdfgh}{e}
Function Structure	{abei}{cdfh}{g}	{abh}{cdefg}{ijk}	{acf}{bdg}{eh}

The observed partition probability analysis method yielded most likely partitions without any overlapping clusters. The results of this analysis clearly show differences in clustering behaviour based on representation modality. For example, comparing most likely partitions for the hair dryer component graph {abegi}{cdfh} with the function graph {abei}{cdfh}{g}, element "g" had a higher

https://doi.org/10.1017/pds.2023.343 Published online by Cambridge University Press

probability of being included with "a", "b", "e", and "i" {abegi} for the component graph and a higher probability of being clustered alone for the function graph {g}. The only two representation comparisons that yielded the same partitioning was the hair dryer's function graph and function structure {abei}{cdfh}{g}. All other comparisons showed differences.

While promising, the results of this analysis method still held some issues worthy of skepticism. The set of observed participant partitions had a wider variety than expected (a larger sample from the complete Bell List of possible partitions). This caused the frequencies of unique observed partitions to be quite low, therefore leading to very small partition probabilities. The method also favors partitions with fewer clusters, which is not desirable. Notably, the toilet's component graph yielded that highest partition probability at 5.62E-04. For the most complex system (the mixer), probabilities were at least an order of magnitude smaller. Given these small probabilities, a third analysis method was implemented.

4.3 Observed variation of information

The third approach relies on information distance to determine the most likely partition. Each partition in the Bell List is compared to partitions generated by the participants. The variation of information metric is used to determine a distance between partitions (Meilă, 2007) as shown in equation (3).

$$VI(X;Y) = -\sum_{ij} r_{ij}\left[\log\left(r_{ij}/p_i\right) + \log\left(r_{ij}/q_j\right)\right] \tag{3}$$

In equation (3), p and q refer to the length subsets in partitions X and Y, respectively. The length of intersection between sets corresponding to p and q is represented by r. Finally, p, q, and r are normalized by the number of elements in the set. For each product-representation pair, an $m \times n$ matrix of distances is generated, where m is the Bell Number for the product and n is the number of participant responses for that product-representation pair. A row-wise mean of the matrix gives the average distance from observed partitions to each partition in the Bell List. The partition corresponding to the minimum distance is then selected as the most likely partition. **Table 3** shows the partitions from the Bell List with the least average variation of information when compared to the observed set of participant partitions.

Table 3: Set partitions with the smallest variation of information

	Hair Dryer	Mixer	Toilet
Component Graph	{abei}{cd}{fg}{h}	{ab}{c}{deh}{fgjk}{i}	{af}{bdg}{c}{eh}
Function Graph	{ai}{b}{cd}{eg}{fh}	{ab}{cde}{fg}{h}{ijk}	{a}{bcf}{d}{eh}{g}
Function Structure	{ae}{b}{cdfh}{g}{i}	{ab}{cd}{eh}{fgjk}{i}	{a}{bc}{d}{eh}{f}{g}

Similar to the observed partition probability method, the most likely partitions do not have overlapping clusters. The variation of information method also shows differences between representation modalities for each of the three systems. For example, the component graph and function graph for the hair dryer yielded Bell List partitions with the least informational differences of {abei}{cd}{fg}{h} and {ai}{b}{cd}{eg}{fh}, respectively. The resulting partitions from this analysis show differences based on representation modality for each of the three systems measured. **Table 4** shows the average information distance for each of the product-representation pairs. They are coded such that green indicates the smallest distance and red indicates the largest distance.

Table 4: Normalized information distance for most common set partitions

	Hair Dryer	Mixer	Toilet
Component Graph	0.290	0.255	0.240
Function Graph	0.292	0.215	0.340
Function Structure	0.264	0.258	0.320

The mixer's function graph {ab}{cde}{fg}{h}{ijk} yielded the smallest information distance across all three systems and representation modalities. This result is somewhat surprising given that the mixer is considered by the research team to be the most complex of the three systems with 11 elements. Spatially, this result makes intuitive sense given that the clusters consist of elements located near each other. **Figure 3** shows the resulting mixer function graph from the variation of information analysis.

https://doi.org/10.1017/pds.2023.343 Published online by Cambridge University Press

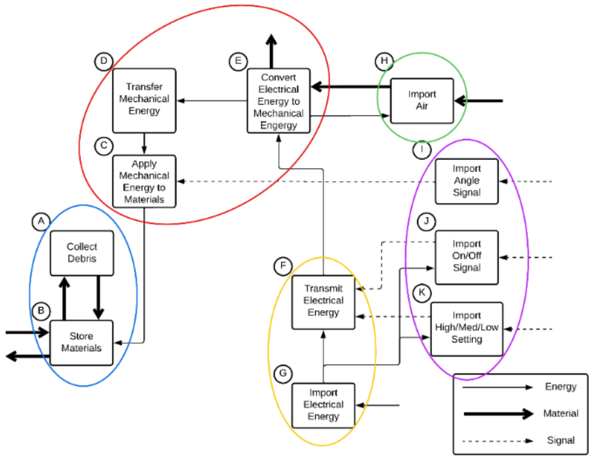

Figure 3. Result for the mixer function graph from the variation of information analysis.

Compared to **Figure 2 (right)** showing the mixer's function graph from element cluster frequency analysis, the partition shown in **Figure 3** shows notable differences using the variation of information method. Notice that the nested clusters in **Figure 2 (right)** have been resolved in **Figure 3** to give a coherent partition of the set. While it might seem that the nested clusters have simply been incorporated into their encompassing cluster, this is not the case for other representations. For example, the toilet function structure showed overlapping clusters of {af} and {ad} using the element cluster frequency analysis. Using the variation of information method, these have been resolved into {a}{d}{f}. In this case, the variation of information method yields a non-overlapping most likely partition that is coherent and accounts for variability in the set of observed participant partitions.

5 DISCUSSION

The results of this work have explored the strengths and weaknesses of three different methods for finding a most likely partition for a subsystem clustering task in an engineering design context. The Element Cluster Frequency method found likely cluster configurations from the set of observed clusters but was unable to parse cases of overlap. The Observed Partition Probability method resolved this issue by focusing on observed partitions instead of observed clusters. However, the frequencies of unique observed partitions were unexpectedly low, therefore leading to small probabilities and some skepticism. Lastly, the Variation of Information method considered informational distance of the observed set of partitions against the entire Bell List. This strategy allowed for the potential discovery of partitions that best match the observed partition set without assuming that the most likely partition had to be within the set of observed participant partitions.

All three analysis methods showed differences in clustering behaviour based on representation modality, which extends the authors' prior work (Murphy *et al.*, 2022; Patel *et al.*, 2022) and partially addresses the first research question. A deeper analysis that investigates specific factors that lead to differences in clustering behavior is left to future work. This has implications for how designers and engineers choose to represent systems, especially during troubleshooting, when dealing with highly complex systems or in high-risk scenarios. In response to the second research question, the Variation of Information method is the most consistent and representative of the observed data. The method gives a most likely partition that does not have overlapping clusters and is not assumed to be within the set of observed partitions. In future work, this could be used to converge on a most likely partition for more complex systems and serves as a foundation to assess other factors that may influence systems thinking.

This work is subject to a few limitations. First, the search for a most likely cluster inherently assumes that a most likely cluster exists. Of course, this may not be the case. Additional analysis and larger sample sizes are needed to investigate whether a most likely cluster theoretically exists. Second, the representations chosen are a small subset of possible representations that might be used to model a system, namely a component graph and functional decompositions. Different designers will create differing functional models for the same product and will also differ in how they spatially arrange the flows and functions. The research presented in this paper suggests that designers should be deliberate when deciding what modality is most appropriate for a given system and context because it impacts

https://doi.org/10.1017/pds.2023.343 Published online by Cambridge University Press

system interpretation, specifically at the subsystem level as shown in the results of this research. Third, engineering components are not necessarily forced into exclusive subsystem clusters outside of the context of this study. This restriction was implemented to highlight the impacts of representation modality. The results would likely be different if an element was allowed to belong to multiple clusters.

6 CONCLUSIONS

The research study presented in this paper has shown differences in modeling behaviour dependent on representation modality through three different analysis methods. The Variation of Information analysis method yielded the most consistent and coherent results during the search for a most likely partition of system elements given designer-generated data. Moving forward, this method will be used for further analysis to investigate which aspects of the representation most strongly relate to differences in observed clustering behavior. In addition, the results of this study are consistent with the authors' prior work that suggested representation modality, spatial arrangement, and subsystem connectivity impact system interpretation. Future work is underway that explores what factors may influence the observed differences in clustering behavior (e.g., spatial arrangement). These results make a significant stride towards understanding how these different factors influence systems thinking.

For complex systems in a hyper-connected world, different modeling behaviours could have a huge impact on how designers interpret, troubleshoot, describe, and understand engineering systems, which is particularly important when dealing with high-stakes engineering scenarios. This work contributes to a larger research endeavour that aims to understand the different aspects of systems thinking and how related factors influence the engineering design process. Exciting future work is planned that compares engineers and designers from different backgrounds, explores how individual differences impact subsystem interpretation, and how systems thinking behaviour relates to mental models of systems.

ACKNOWLEDGMENTS

This work is supported by the Engineering Fellows Postdoctoral Fellowship Program through grant number 2127509. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or the American Society of Engineering Education.

REFERENCES

Abbott, D. and Lough, K.G. (2007) ‘Component Functional Templates as an Engineering Design Teaching Aid’, in 2007 *ASEE Annual Conference & Exposition*, pp. 12–388.

Barlow, T. *et al.* (2021) ‘A System Design Optimization Model for Integrated Natural Resource Conservation and Development in an Agricultural Community’, *Proceedings of the Design Society*, 1, pp. 273–282. Available at: https://doi.org/10.1017/pds.2021.28.

Buede, D.M. and Miller, W.D. (2016) ‘The Engineering Design of Systems: Models and Methods’.

Dieter, G.E. and Schmidt, L.C. (2009) *Engineering Design*. McGraw-Hill Higher Education Boston.

Erden, M.S. *et al.* (2008) ‘A Review of Function Modeling: Approaches and Applications’, *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 22(2), pp. 147–169. Available at: https://doi.org/10.1017/S0890060408000103.

Forte, S., Göbel, J.C. and Dickopf, T. (2021) ‘System of Systems Lifecycle Engineering Approach Integrating Smart Product and Service Ecosystems’, *Proceedings of the Design Society*, 1, pp. 2911–2920. Available at: https://doi.org/10.1017/pds.2021.552.

Friedenthal, S., Alan, M. and Steiner, R. (2008) *Practical Guide to SysML*. Elsevier. Available at: https://doi.org/10.1016/B978-0-12-374379-4.X0001-X.

Gardner, M. (1978) ‘Bells-Versatile Numbers that can Count Partitions of a Set, Primes and Even Rhymes’, *Scientific American*, 238(5), pp. 24-+.

Halmos, P. (1998) *Naive Set Theory (Undergraduate Texts in Mathematics)*. New York: Springer.

Hirtz, J. *et al.* (2002) ‘A Functional Basis for Engineering Design: Reconciling and Evolving Previous Efforts’, *Research in engineering Design*, 13(2), pp. 65–82.

Kai, L.T. (1997) ‘Bell Numbers and Bell Numbers Modulo a Prime Number’, *Mathematical Medley*, 42(2), pp. 55–58.

Liem, A. (2017) ‘Extending System Design Tools to Incorporate User-and Contextual Elements in Developing Future Products and Services’, in *DS 87-3 Proceedings of the 21st International Conference on Engineering Design (ICED 17) Vol 3: Product, Services and Systems Design, Vancouver, Canada, 21-25.08.* 2017, pp. 161–170.

Meilă, M. (2007) ‘Comparing Clusterings—An Information Based Distance’, *Journal of Multivariate Analysis*, 98(5), pp. 873–895. Available at: https://doi.org/10.1016/j.jmva.2006.11.013.

Murphy, A.R. *et al.* (2019) ‘Graduate Students’ Mental Models: An Investigation Into the Role of Function in Systems Understanding’, in *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. American Society of Mechanical Engineers, p. V007T06A024.

Murphy, A.R. *et al.* (2020) ‘Validation of a Mental Model Elicitation Instrument through Deployment of Control Groups in an Undergraduate Engineering Program’, in 2020 *ASEE Virtual Annual Conference Content Access*.

Murphy, A.R. (2021) *Leveraging Design Knowledge: Strategies to Improve Systems Understanding*. Georgia Institute of Technology.

Murphy, A.R. *et al.* (2022) ‘The Effects of Representation Modality on Subsystem Clustering’, in *ASME 2022 International Design Engineering Technical Conferences*. St. Louis, MO: ASME.

Nagel, R.L. *et al.* (2015) ‘Improving Students’ Functional Modeling Skills: A Modeling Approach and a Scoring Rubric’, *Journal of Mechanical Design*, 137(5), p. 051102.

Otto, K.N. and Wood, O.L. (2003) *Product Design: Techniques in Reverse Engineering and New Product Development*. 1st edn. Upper Saddle River Nj: Prentice Hall.

Pahl, G. *et al.* (2007) ‘Engineering Design: A Systematic Approach Third Edition’, *Berlin, Springer Science+ Business Media Deutschland GmbH, 2007. 632* [Preprint].

Patel, A. *et al.* (2022) ‘Role of Representation in Subsystem Clustering: Effects of Distance Between Elements’, in *ASME 2022 International Design and Technical Conferences*. St. Louis, MO: ASME.

Patel, A. and Summers, J.D. (2021) ‘Exploring the Effects of Individual Differences in Function Structure Modeling Behaviors’, in *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. American Society of Mechanical Engineers, p. V006T06A036.

Patrikainen, A. and Meila, M. (2006) ‘Comparing Subspace Clusterings’, *IEEE Transactions on Knowledge and Data Engineering*, 18(7), pp. 902–916. Available at: https://doi.org/10.1109/TKDE.2006.106.

Piesk, T. (2011) *File: Set Partitions 5; Circles.svg*, *Wikipedia Commons*.

Qian, L. and Gero, J.S. (1996) ‘Function–Behavior–Structure Paths and Their Role in Analogy-Based Design’, *AI EDAM*, 10(4), pp. 289–312.

Rahman, M.H. *et al.* (2018) ‘Automatic Clustering of Sequential Design Behaviors’, in *Volume 1B: 38th Computers and Information in Engineering Conference*. American Society of Mechanical Engineers, pp. 1–14. Available at: https://doi.org/10.1115/DETC2018-86300.

Rossi, G. (2011) ‘Partition Distances’, pp. 1–22. Available at: http://arxiv.org/abs/1106.4579.

Sanfeliu, A. and Fu, K.-S. (1983) ‘A Distance Measure Between Attributed Relational Graphs for Pattern Recognition’, *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13(3), pp. 353–362. Available at: https://doi.org/10.1109/TSMC.1983.6313167.

Shai, O. and Preiss, K. (1999) ‘Graph Theory Representations of Engineering Systems and their Embedded Knowledge’, *Artificial Intelligence in Engineering*, 13(3), pp. 273–285. Available at: https://doi.org/10.1016/S0954-1810(99)00002-3.

Stauffer, M. *et al*. (2017) ‘A Survey on Applications of Bipartite Graph Edit Distance’, in P. Foggia, C.-L. Liu, and M. Vento (eds) *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Cham: Springer International Publishing (Lecture Notes in Computer Science), pp. 242–252. Available at: https://doi.org/10.1007/978-3-319-58961-9_22.

Stone, R.B. and Wood, K.L. (1999) ‘Development of a Functional Basis for Design’, in *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. American Society of Mechanical Engineers, pp. 261–275.

Summers, J.D., Eckert, C. and Goel, A.K. (2017) ‘Function in Engineering: Benchmarking Representations and Models’, *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 31(4), pp. 401–412. Available at: https://doi.org/10.1017/S0890060417000476.

Wagner, R.A. and Fischer, M.J. (1974) ‘The String-to-String Correction Problem’, *Journal of the ACM*, 21(1), pp. 168–173. Available at: https://doi.org/10.1145/321796.321811.

Waschull, S. *et al.* (2020) ‘Work Design in Future Industrial Production: Transforming Towards Cyber-Physical Systems’, *Computers & Industrial Engineering*, 139, p. 105679. Available at: https://doi.org/10.1016/j.cie.2019.01.053.

Wichmann, R.L., Eisenbart, B. and Gericke, K. (2019) ‘The Direction of Industry: A Literature Review on Industry 4.0’, *Proceedings of the Design Society: International Conference on Engineering Design*, 1(1), pp. 2129–2138. Available at: https://doi.org/10.1017/dsi.2019.219.

Zeng, J. *et al.* (2020) ‘A Survey: Cyber-Physical-Social Systems and their System-Level Design Methodology’, *Future Generation Computer Systems*, 105, pp. 1028–1042. Available at: https://doi.org/10.1016/j.future.2016.06.034.

https://doi.org/10.1017/pds.2023.343 Published online by Cambridge University Press