Proceedings of the Design Society, Volume 5: ICED25
https://doi.org/10.1017/pds.2025.10301

Semantic and network analysis of design protocols to reveal design framing phenomena

Nick Kelly [1]✉ **and John S. Gero** [2]

[1] *Queensland University of Technology, USA,* [2] *University of North Carolina at Charlotte, USA*

✉ nick.kelly@qut.edu.au

ABSTRACT: This paper introduces a novel approach to analysing design protocols using a combination of methods. It describes an approach using a synthesis of concept extraction (using an LLM), semantic analysis (using word vectors and conceptual clustering), and network analysis (following graph construction). It suggests that the resulting measures are useful for studying design framing and for aiding qualitative analysis. It demonstrates this technique with data from a study of 17 designers addressing two design problems. The method enables the comparison of designers working on the same problem as well as the study of individual designers' use of concepts over time during a think-aloud study.

KEYWORDS: design cognition, research methodologies and methods, computational design methods, design framing, semantic data processing

1. Introduction and background

Design is a goal-directed activity (Gero, 1990; Simon, 1969). It involves designers taking actions directed towards an internal conception of the design goals. The term *design frame* is used to refer to this internal conception that a designer has for what they are doing. A design frame involves a way of understanding the problem, a way of working, and a sense of the value being aimed for (Dorst, 2011). In this paper, a design frame is taken to be a conceptual assemblage that does these things; an internal structure composed of concepts (Kelly & Gero, 2022). This definition follows an analysis of uses of the term design frame in the literature revealing multiple and contradictory uses (Kelly & Gero, 2022).

A core question in design research, at least since the time of Schön (1983), has been: How do designers construct and change their understanding of a design problem and its context throughout design activity? Satisfactory answers remain elusive at the level of cognitive processes. Yet there is broad agreement on the importance of understanding the cognitive and metacognitive abilities of designers, that is, the way that they work with concepts and understand their own ways of working with concepts (Ball & Christensen, 2019; Hay et al., 2017). Advances in methods to analyse design protocols for design framing phenomena help to address fundamental questions of why and how frames change. Such questions are significant for advancing scientific understanding of what designers do (design science) and how people gain capabilities for design (design education). For example, how do explicit instruction about approaches to design framing or the provision of real-time feedback change how somebody designs (Kelly et al., 2024)? This paper contributes to methods that can, in future, be used to respond to such important questions.

The methods of inquiry and the language to describe design framing have evolved through decades of research. A widely used method is to code design protocols for episodes of naming and framing and then identify the moving and reflecting that occurs within a frame, Figure 1a (Valkenburg & Dorst, 1998). This method is useful to qualitatively understand how a design frame shifts over time, to make visible the conceptual development of a designer or design team. Another method that makes links between concepts salient is linkography, showing how the concepts developed over time are connected to one

https://doi.org/10.1017/pds.2025.10301 Published online by Cambridge University Press

another (Goldschmidt, 2014). Linkography permits quantitative analysis of the emergent conceptual lattice as it develops over time and provides the potential to see the genesis of ideas within the arc of a design problem by looking at forward and backward links, Figure 1. Casakin et al. (2024) introduced a data-driven approach to quantitatively characterize and measure framing and reframing behaviors during design problem-solving, using the concept of "first occurrence" of ideas in a design session (Gero & Kan, 2016).

This paper introduces and demonstrates a novel technique for bringing together semantic analysis through hierarchical word-vector embeddings and network analysis. It uses network and cosine similarity measures to permit comparison of designers addressing the same design task as well as analysis of individual designers addressing a design problem over time. It uses extraction of concepts from design protocols and similarity measurements to permit low-resource and high-precision qualitative analysis of designing. It builds on the work of Kelly et al. (2024) who developed methods for graphing design concepts over time. The method demonstrated here contributes to the agenda laid out by Kelly & Gero, of investigating how designers move around within their own knowledge, Figure 1c (2015). This paper makes a contribution by describing the novel method that blends existing approaches of automated concept identification (with LLMs), contextual word vector embeddings, clustering, and network analysis (graph construction and measurement).

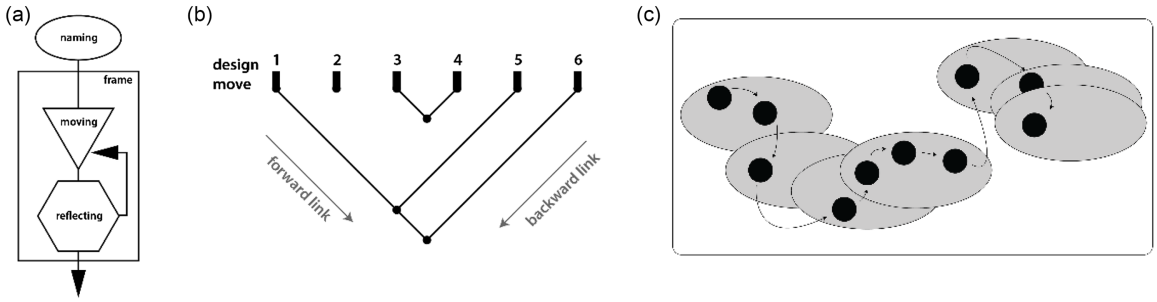


Figure 1. Three different approaches to understanding designers' cognition during design (a) naming-framing-moving-reflecting method of (after Valkenburg & Dorst, 1998); (b) the linkography method (after Goldschmidt, 2014); and (c) design activity conceived as a shifting conceptual assemblage within a designer's own knowledge as inspiration for this paper (Kelly & Gero, 2015)

2. Methods

The method will be explained by first introducing two design tasks and the participants in a design protocol study. The study is part of a broader US NSF-funded experiment producing fMRI data looking at fixation (Kim et al., 2024). This section then describes the methods for semantic and network analysis of design protocols that are the focus of this paper. Results are then included as a demonstration of how this method permits new insights from design protocols into design framing phenomena, i.e., the dynamics of evolving conceptual assemblages that constitute a significant aspect of design cognition.

2.1. Design problems and participants

The experiment from which this verbalization data is drawn involves concurrent multi-modal measurements using fMRI for neurocognitive activity, think-aloud for cognitive activity and sketches for final design activity (Kim et al., 2024). Participants were given two tasks. The first task was the design of a spill-proof coffee cup. Instructions to participants read as follows: "You are asked to construct an inexpensive, disposable, spill-proof coffee cup. You should construct as many designs as possible, write comments with each design, and number each individual design. There are no constraints in the materials you may want to use. The problems to be addressed are: (1) Leaking of the cup if it tips over; (2) Leaking of the cup when squeezed; and (3) Hot liquid burning the user's mouth."

The second task was the design of a new kind of bike rack for a car. Instructions to participants included Figure 2 and the following text: "This is an example of a present-day bike rack. It is a top-mount design having suction cups holding it to the car roof and railings for the bicycle tires. The bicycle is set in the rails and the vinyl coated hook is attached to the seat tube of the bike and then the hook is tightened down by hand with a wing nut. The problem in this case is the difficulty of mounting the middle bikes on the

https://doi.org/10.1017/pds.2025.10301 Published online by Cambridge University Press

rack. Suppose you are asked to construct a new bike rack for cars. You should construct as many designs as possible, write comments with each design, and number each individual design. There are no constraints in the materials you may want to use. The problems to be addressed are: (1) Easy mounting of the bicycle; (2) Easy mounting of the rack; (3) Cannot harm bike or car; and (4) Must be versatile for all bikes and cars."

The study reported in this paper involved data from the first 17 participants of this larger study for whom we had complete datasets (i.e., protocols from both problems). Participants were volunteers, recruited from freshman (first year) students in the mechanical engineering degree in a private US east coast university under the university's IRB arrangement, Participants were paid cash as an incentive to participate. All design activity occurred inside an fMRI machine and participants were able to sketch while designing. This setting and the sketching are irrelevant to the methods that are the focus of this paper but are significant for interpreting the results included here. Designing continued for 16 minutes. The think aloud recordings were transcribed, cleaned, and divided into 2-second segments to match the TR (repetition time) in the fMRI scanning. This transcript was the input for the process described.

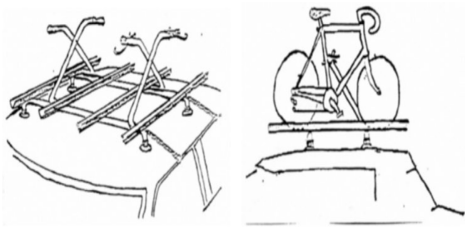


Figure 2. Image included alongside instructions in bike rack design problem (Kim et al., 2024)

The pipeline for the entire method is shown in schematic form in Figure 3. Each of these steps is then described in more detail.

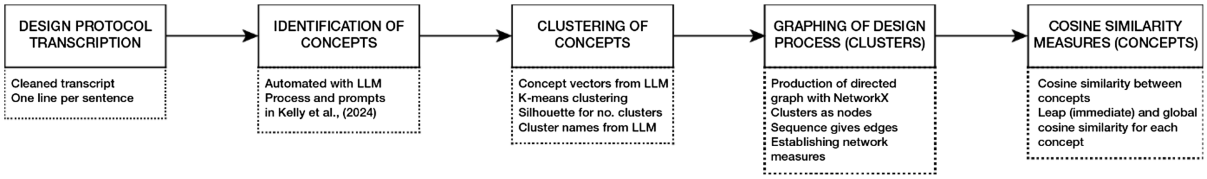


Figure 3 Pipeline for analysis of design protocol, from input of transcript to output of network and semantic measures

2.2. Coding concepts

Prior work has shown human coding of concepts and suggested the possibility of automating this with LLMs (Kelly et al., 2024). In this work we used a static LLM from OpenAI, "gpt-4o-2024-08-06", accessed through API. Coding of concepts within transcripts was carried out by the LLM using the method and prompts described in this prior work by Kelly et al. (2024) with conceptual underpinning in situated cognition (Barsalou, 2005). The essence is that a concept is a noun along with its modifiers and adjectives but not pronouns or verbs.

Using an LLM for identifying concepts is inherently unreliable, both internally (repeated coding of the same transcript) and when compared to humans. This is a significant limitation of the method that is counterbalanced by the speed and low resources required when compared to human coding of concepts, though further research would be required to quantify this trade-off.

2.3. Semantic distance

Vectors within semantic word space were established for each identified concept. Word embeddings came from OpenAI's model "text-embedding-3-small". Weightings in establishing vectors were 0.8 for

https://doi.org/10.1017/pds.2025.10301 Published online by Cambridge University Press

the full concept and 0.2 for the full sentence (including the concept) to shift weightings based on context of concept usage, i.e., the same concept will have different weights when used within different sentences. Similarity between any two concepts was calculated as the cosine similarity between vectors, where a higher cosine similarity indicates more similarity between concepts (Salton, 1983). Two cosine similarity measures were used in the study. The first is *leap similarity*, the similarity of the present concept to the immediate previous concept used by the designer. The second is *global similarity*, the similarity of the present concept to the most similar concept from all concepts that came previously. By definition, global similarity will always be equal to or greater than leap similarity.

Concept clustering

Concepts were grouped into clusters based upon their vectors. The optimal number of clusters was determined using K-means clustering and the silhouette method for establishing optimal K (max. 50). A critique of the work of Kelly et al. (2024) is that they treat concepts that are similar as entirely different entities. The motivation for clustering is to have graphs represent movement between regions in conceptual space. Word vector space is taken as a proxy for conceptual space, following the work of Gärdenfors (2004) and others. Figure 4 shows an example of the clusters produced from a set of concepts for one participant, using t-SNE dimensional reduction to permit graphing of a high dimension space in two dimensions (axes have meaningless labels for this reason).

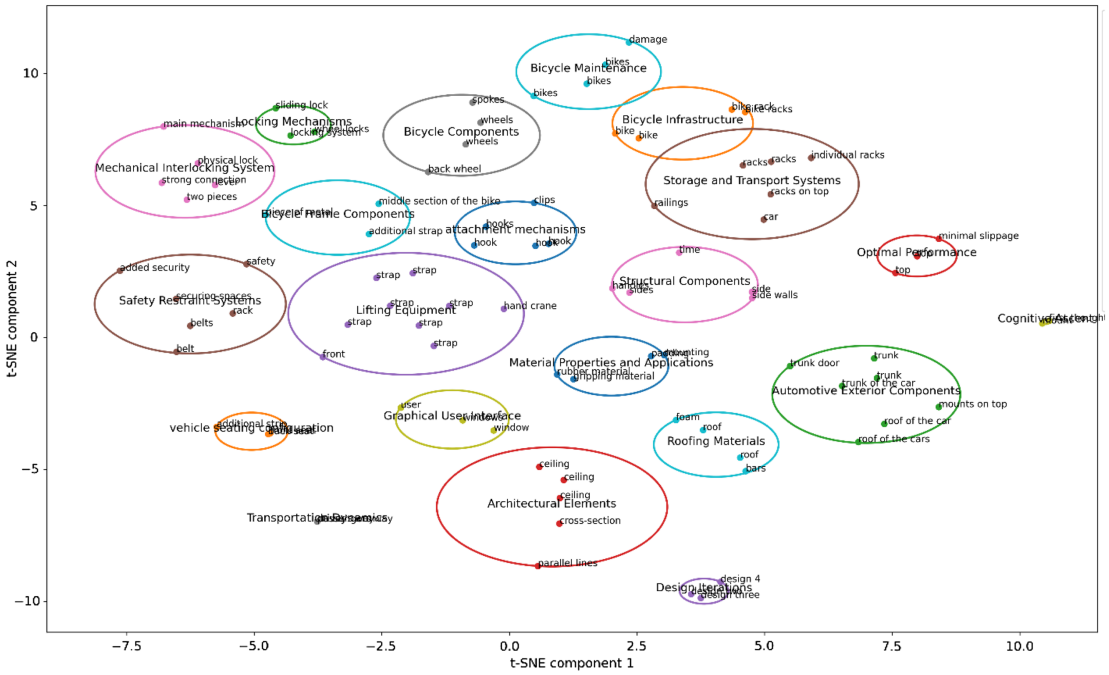


Figure 4. Clustering of concepts for a participant in the bike problem results in identification of regions of conceptual space. Here, clusters have been given names by ChatGPT and high dimensional word vector space has been projected onto two dimensions using t-distributed Stochastic Neighbour Embedding (t-SNE)

2.4. Network measures

The concepts detected within a protocol were then used to construct a graph. Each region of conceptual space, which contains a cluster of concepts used by the designer, becomes a node/vertex within the graph. Each time a concept is found within the transcript, a directed edge $R_A \rightarrow R_B$ is drawn between the region R containing concept A and that containing concept B. The resulting directed graph $G = (V,E)$ represents the design protocol as a sequence of conceptual regions (vertices) V and the connections between them (edges) E. Some edges are self-connecting for multiple sequential concepts within the same region. Within the resulting graph a range of network measures were utilised, specifically: numbers of nodes and edges, degree, connectivity, longest-shortest path, and clustering (Wasserman, 1994).

https://doi.org/10.1017/pds.2025.10301 Published online by Cambridge University Press

3. Results

This section includes two sets of results. The method described here permits standardised, quantitative comparison of designers within a single problem. This approach enables inferences to be made about the different cognitive styles of individual designers across problems. We will refer to this as the cognitive style of the designer, following Herb Simon, who noted that "those differences among designs that we call 'style' arise from a variety of sources, and that the design process is one of the most important of these" (Simon, 1975, p. 10). For example, it provides a basis for making claims about the transfer of framing style for a designer across different design problems, which would have significance for design education. It also permits comparison of cognitive style between designers within each problem, to establish a range of different approaches to the same problem. It is claimed that the proposed method is less resource intensive than either framing analysis or linkography and permits novel kinds of analysis. This results set is intended to demonstrate the utility of this novel method.

3.1. Comparing designers across both problems

For comparing designers' response to each problem, the concepts from all 17 designers were pooled prior to clustering. The clusters that emerged through K-means clustering (e.g., Figure 4) were treated as regions of conceptual space; all concepts within a region are assumed to have a meaningful semantic relationship, based on cosine similarity, and the graph that is created represents movement between these regions of conceptual space. This is a form of hierarchical clustering, with clusters as an abstraction from the concepts within them. It prevents similar concepts (e.g., 'wheel' and 'bike wheel') from being treated as distinct entities when graphing: it is desirable that the structure of the graph represent movement through conceptual space, revealing revisitation of past spaces (like a Linkograph) rather having each distinct concept be treated as a node (a limitation in the work of (Kelly et al., 2024)).

This technique permitted a comparison of the way that each designer navigated through the same regions of design space, after partitioning it into these emergent conceptual regions. Tables 1 and 2 show participants' eight network metrics ordered by participant number. Three measures that we considered as interesting for revealing design process styles are *average clustering, average leap similarity* and *average global cosine similarity*.

Each node in a graph has a clustering coefficient, the number of connected triads of possible connected triads. Average clustering is the average of these values for all nodes. We suggest that high average clustering might indicate a style for moving frequently between clusters; perhaps trying on new ideas in rapid succession. The correlations in Table 3 show that clustering changed between the two problems (i.e., was not consistent for each designer).

Average leap similarity is an indicator of how semantically distant a designer tends to move in relation to each prior concept. Average global similarity is a measure of how semantically distant each new concept is in relation to all that came before it. In neither case was there a strong correlation for the same designer across problems.

Table 1. Network measures for comparing participants in the bicycle problem in the same design space. Participant ID, number of nodes, number of edges, avg. degree, avg. connectivity, longest shortest path, avg. clustering, avg leap (avg. cosine similarity between sequential concepts), avg. global (avg. cosine similarity between concept and most similar prior concept)

PID	Nodes	Edges	Avg. degree	Avg. connectivity	LSP	Avg. clustering	Avg. leap	Avg global
1	22	50	4.55	1.33	7	0.22	0.52	0.73
3	32	101	6.31	1.81	8	0.15	0.53	0.73
5	31	105	6.77	1.97	7	0.12	0.49	0.72
6	23	57	4.96	1.55	9	0.14	0.48	0.74
7	31	78	5.03	1.51	9	0.08	0.45	0.69
8	34	82	4.82	1.36	10	0.09	0.49	0.73
9	36	133	7.39	2.41	6	0.14	0.47	0.68
10	26	91	7.00	1.87	6	0.23	0.49	0.77
11	38	137	7.21	2.17	6	0.11	0.46	0.72

(*Continued*)

https://doi.org/10.1017/pds.2025.10301 Published online by Cambridge University Press

Table 1. Continued.

PID	Nodes	Edges	Avg. degree	Avg. connectivity	LSP	Avg. clustering	Avg. leap	Avg global
12	39	160	8.21	2.26	8	0.17	0.47	0.66
13	21	67	6.38	1.66	6	0.28	0.48	0.71
14	33	103	6.24	1.57	6	0.26	0.39	0.69
15	37	116	6.27	1.71	7	0.11	0.43	0.71
16	29	89	6.14	1.47	7	0.29	0.44	0.77
17	24	67	5.58	1.49	9	0.07	0.50	0.75
18	35	115	6.57	1.67	9	0.20	0.47	0.69
19	33	111	6.73	1.94	7	0.21	0.43	0.72
Avg	30.82	97.76	6.24	1.75	7.47	0.17	0.47	0.72

Table 2. Network measures for comparing participants in the coffee cup problem in the same design space. Participant ID, number of nodes, number of edges, avg. degree, avg. connectivity, longest shortest path, avg. clustering, avg leap (avg. cosine similarity between sequential concepts), avg. global (avg. cosine similarity between concept and most similar prior concept)

PID	Nodes	Edges	Avg. degree	Avg. connectivity	LSP	Avg. clustering	Avg. leap	Avg global
1	17	28	3.29	0.91	8	0.05	0.46	0.68
3	26	69	5.31	1.26	7	0.11	0.53	0.73
5	28	81	5.79	1.50	6	0.19	0.45	0.68
6	26	54	4.15	1.34	9	0.16	0.35	0.71
7	31	68	4.39	1.14	8	0.08	0.48	0.67
8	32	82	5.13	1.48	9	0.21	0.45	0.66
9	27	90	6.67	1.82	6	0.18	0.44	0.67
10	28	82	5.86	1.83	7	0.11	0.33	0.69
11	32	118	7.38	2.36	7	0.15	0.40	0.70
12	32	116	7.25	2.01	7	0.16	0.45	0.70
13	23	55	4.78	1.14	7	0.24	0.43	0.59
14	30	88	5.87	1.56	7	0.16	0.43	0.69
15	28	71	5.07	1.30	9	0.14	0.40	0.66
16	29	94	6.48	1.64	7	0.26	0.39	0.70
17	21	62	5.90	1.60	7	0.21	0.39	0.69
18	29	82	5.66	1.55	7	0.18	0.42	0.72
19	33	95	5.76	1.57	7	0.15	0.40	0.73
Avg	27.76	78.53	5.57	1.53	7.35	0.16	0.42	0.69

Network measures that are closely related to the total number of concepts (nodes, edges, degree) were highly correlated across problems. This indicates that participants are using a similar number of concepts in each problem, i.e., closely related to them talking at a similar rate in each problem.

Connectivity, a function of how many nodes must be removed from a network to disconnect it, correlates significantly ($p<0.005$) between the two problems. This indicates that the robustness of the graphs (is similar across the two problems. Degree, the ratio of nodes to edges, correlates significantly with

Table 3. Correlations between each network measures across the two design problems (coffee to bike) for all 17 participants

Measure	Correlation (coffee to bike)
Nodes	0.75
Edges	0.86
Avg. degree	0.83

(*Continued*)

https://doi.org/10.1017/pds.2025.10301 Published online by Cambridge University Press

Table 3. Continued.

Measure	Correlation (coffee to bike)
Avg. connectivity	0.68
Longest shortest path	0.51
Avg. clustering	0.19
Avg. leap similarity	0.32
Avg. global similarity	0.13

connectivity in both directions: bike degree to coffee connectivity (0.721) and coffee degree to bike connectivity (0.710). It is possible that this result of connectivity being correlated is a function of similar numbers of nodes and edges across problems.

These measures make it possible to delineate empirically between different designers addressing the same problem in a standardised way; with the recognition that exactly what is being delineated as a construct has yet to be established.

3.2. Selecting designers for further analysis

We ranked designers by average global similarity across the two problems. This measure involved noting the global similarity for each concept when it occurred within the transcript and then computing the average and then ranking designers for each problem. This permitted us to select the designer with the 'highest novelty in leaps' across both problems (lowest average global cosine similarity, a proxy for most novelty in conceptual leaps) as participant P9; and the designer with the 'lowest novelty in leaps' (i.e., highest average global cosine similarity) as participant P3. The hypothesis is that this measure can reveal differences in cognitive styles of these designer.

The method permits analysis of how an individual designer's process develops over time. Figures 5 and 6 show graphs of changing similarity measures for participant P9. The two measurements of *leap similarity* and *global similarity* change over time and can be used to identify two phenomena.

Firstly, divergence between global and leap cosine similarity is indicated by a gap in the two lines, such as at concept 94 in Figure 5 or 106 in Figure 6. These points of divergence occur when the global

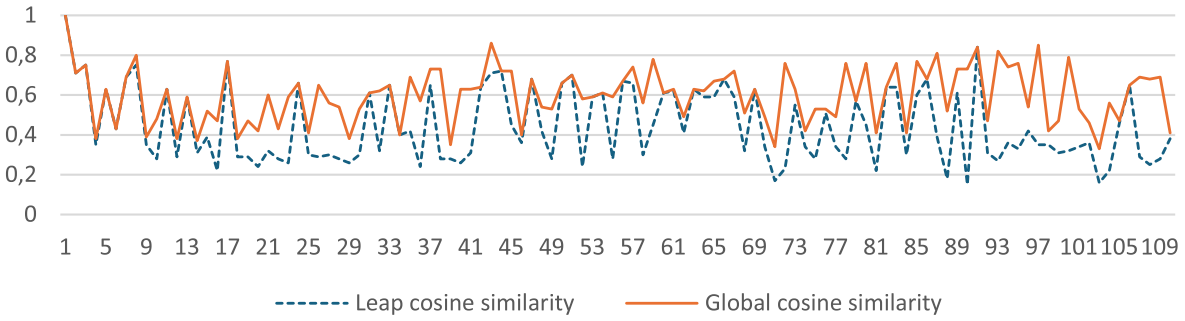


Figure 5. Global and immediate cosine similarity measures over time for participant P9 in the coffee problem (x-axis is enumeration of concepts in order of occurrence)

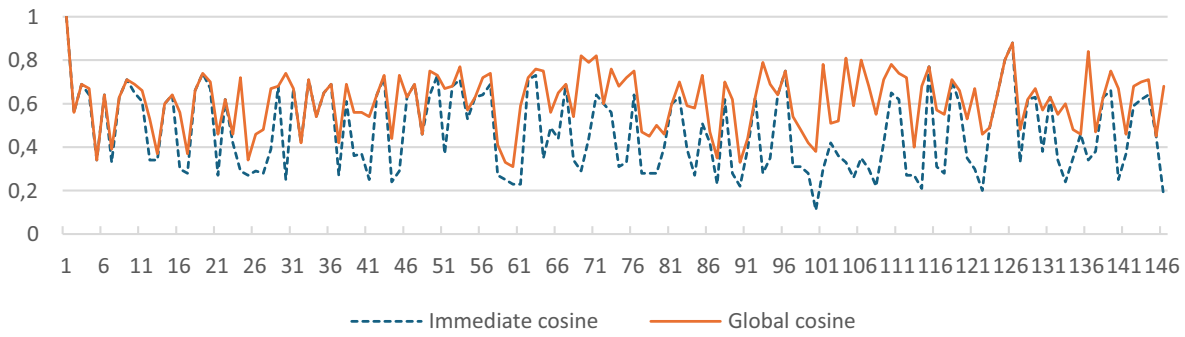


Figure 6. Global and immediate cosine similarity measures over time for participant P9 in the bike problem (x-axis is enumeration of concepts in order of occurrence)

https://doi.org/10.1017/pds.2025.10301 Published online by Cambridge University Press

similarity is high and the leap similarity is low indicating that the designer is revisiting an old concept (high global) but has arrived here from a distant part of conceptual space (low leap). Such moments can then be associated with the moments in the transcript for analysing design framing phenomena; in the latter case the designer has returned to consideration of the 'user' after a passage of design considering how hooks and padding on the bike rack might be structured.

These results can be compared across different designers. Figures 7 and 8 are graphs for P3 who, as seen in Tables 1 and 2, had higher average global cosine similarity (more similar concepts with each move, in general) when compared to P9. In these graphs over time we can observe that global similarity for P9 (Figures 5 & 6) regularly goes below 0.5 throughout the design process for both problems (30 and 35 times respectively). In contrast, P3 (Figure 7 & 8) rarely has global similarity drop below 0.5 for either problem (12 and 16 times respectively). These graphs show time-based details, such as the rhythms with which more novel concepts occur: spikes of low global similarity, particularly those occurring later within the design activity. Figure 9 shows the networks for the two designers for the coffee problem graphed in the same 2D vector space with the same conceptual regions. P3 visits fewer regions.

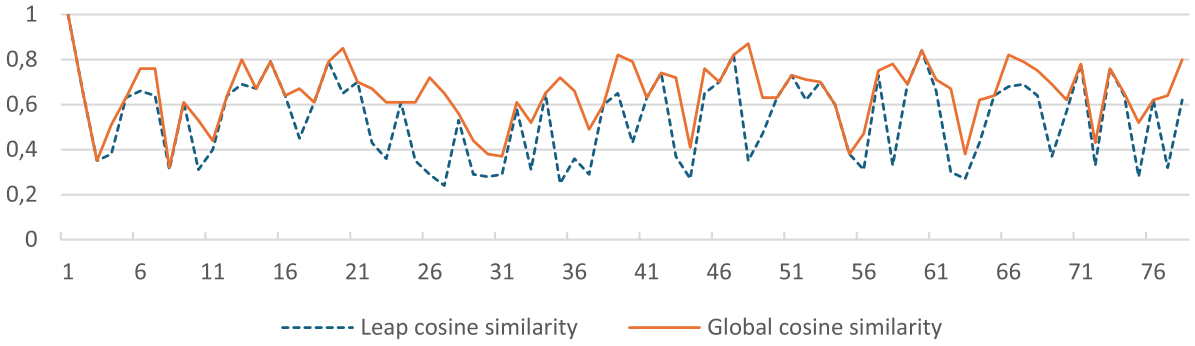


Figure 7. Global and immediate cosine similarity measures over time for participant P3 in the coffee problem (x-axis is enumeration of concepts in order of occurrence)

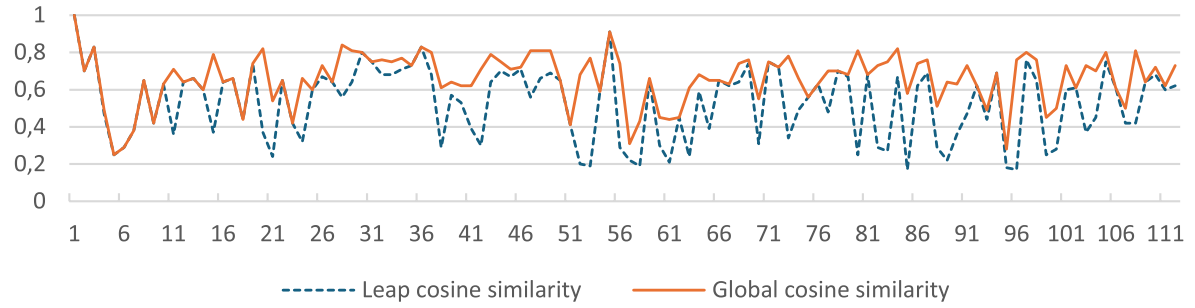


Figure 8. Global and immediate cosine similarity measures over time for participant P3 in the bike problem (x-axis is enumeration of concepts in order of occurrence)

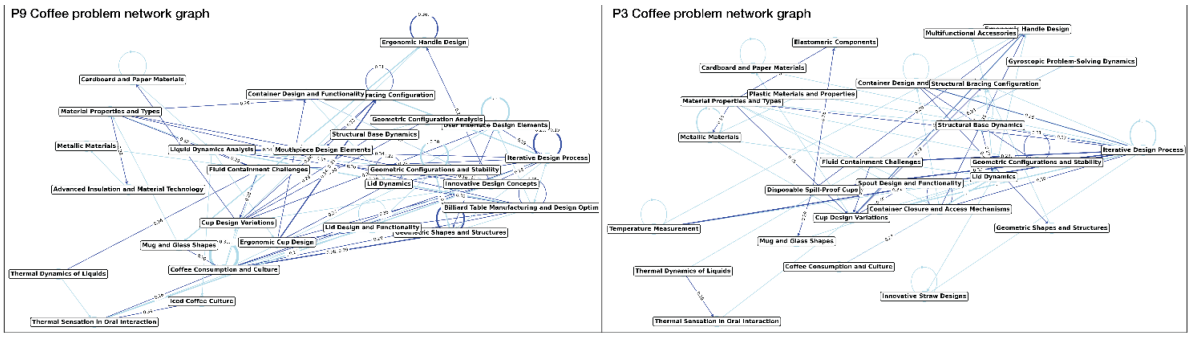


Figure 9. Comparison of network graphs for P9 and P3. Labels indicate nodes and positions of conceptual regions are consistent across the two graphs (within 2D vector space)

https://doi.org/10.1017/pds.2025.10301 Published online by Cambridge University Press

The method provides an output of a table of concepts in the order that they appear, with similarity measures at each step. This permits precise identification of phenomena (using empirical measures) to be associated with qualitative analysis (within tables of concepts), with reference to the original transcript available where required.

To demonstrate this utility of the method we developed a way of identifying significant design framing phenomena by finding concepts that have: (1) the lowest global cosine similarity values (as proxy for most novel); where those global minima (2) occur after the first 10 concepts, because global similarity becomes more interesting after the initial passage of design (during which it follows leap similarity); and (3) that were on task for design rather than dialogue about the experiment (e.g., "where is the eraser?"), as dialogue about the experiment was coded as quite novel perhaps as it was related to a different goal. The results for the top three concepts for each designer in the bike rack problem are shown in Table 4 along with the relevant passage of design. In each case we propose that these are interesting framing phenomena, bringing into the design structural considerations (e.g., the new material of titanium or the new structural relationship of a waffle pattern), behavioral considerations (e.g., bumps affecting performance of design) and functional considerations (e.g., the need for the rack to work on an SUV as well as a car).

Table 4. Specific concepts as identified through global similarity with extracts from the transcript

PID	Concept (order)	Global	Extract from transcript to provide context
P9	titanium (89)	0.33	Then there are other things to address that are not in the design of the bike rack, but this can just be any kind of thicker . . . Let's even use titanium for this example. If we're not limited in materials, something very durable and lightweight, uh and then we would just have a padded end right here and this would go . . .
P9	serious enough bumps (86)	0.35	Looking at this part, this does not need to hook into the trunk if you're hitting serious enough bumps to have this much weight go flying then there are other things to address that are not in the design of the bike rack
P9	SUV	0.37	If you are putting your bikes on an SUV, you might be having to lift a bike maybe seven feet in the air to get on top of the SUV. However, with this approach of putting it on the back of the car a person of any height or strength . . . could at least get it a few feet off the ground to be able to mount to the trunk.
P3	waffle pattern action	0.31	So a box, right? We do some waffle pattern action. And it's just like, like that, but like more organized right, like that like more organized right but we just put hooks everywhere like every little spot has a hook right?
P3	top mount	0.38	If it has to be a top mount maybe maybe maybe we could do like a magnet looking thing.
P3	aluminum	0.41	I don't want to say iron because I feel like that would be too heavy to put on . . . Leave you go with like, what's lighter, aluminum?

4. Discussion and conclusion

This paper has demonstrated novel ways to use network and cosine similarity measures for analysing design protocols. The method appears to have utility for studying design framing, though this utility has only been demonstrated in a limited way with 17 participants addressing two problems (network measures) and two of those participants (more detailed analysis).

The first significant idea within the paper is that clustering concepts into regions, based on semantic similarity, prior to graphing, gives the resulting graph a structure that is more significant for investigating cognitive behaviours of designers, such as framing. In this way it implements past ideas for investigating design framing phenomena (Gardenfors, 2004; Kelly & Gero, 2015) and addresses limitations in recent work (Kelly et al., 2024). It introduces a novel approach for visualising and measuring aspects of framing: the way that designers bring together a conceptual assemblage in response to a problem and shift it through design activity (Kelly & Gero, 2022).

The second is that partitioning a semantic space into conceptual regions, and then graphing the design process for multiple designers on that partitioned space, provides a new way of meaningfully comparing

https://doi.org/10.1017/pds.2025.10301 Published online by Cambridge University Press

cognitive design styles. This idea has only been demonstrated here as more theorising and experimentation is required to determine the meaning and reliability of these measures.
The third significant idea is that producing tables for sequential conceptual development within a problem with associated measures can be of use in identifying where within a protocol transcript significant design framing phenomena are occurring. It has developed two heuristics that are shown to be useful for identifying such phenomena. These are that gaps between local and global similarity indicate revisiting old concepts, and that certain global similarity measures indicate significant new ideas. These proposed measures require further investigation for understanding their validity.
The fourth is that this entire process was automated using an LLM making it possibly to rapidly analyse hundreds or thousands of design protocols in this way. Our future work will look at the design of experiments that can use such automated analysis to look for convergence of findings across many different protocol studies using a dataset that we have begun to assemble. The focus of the work remains on understanding how designers develop and change their framing for a design problem.

Acknowledgments

This work is supported in part by a grant to J.S.G. from the NSF grant no. DRL-2100138. N.K. gratefully acknowledges the roles of Jess Greentree, Ricardo Sosa, and Richard Evans for their contributions to the work underpinning the methods described in this paper.

References

Ball, L. J., & Christensen, B. T. (2019). Advancing an understanding of design cognition and design metacognition: Progress and prospects. *Design Studies*, 65, 35–59. https://doi.org/10.1016/j.destud.2019.10.003
Barsalou, L. W. (2005). Situated conceptualization. In *Handbook of categorization in cognitive science* (pp. 619–650). Elsevier.
Casakin, H., Sopher, H., Gero, J. S., & Anidjar, O. H. (2024). The use of occurrences of ideas for constructing and characterizing the design space. *Proceedings of the Design Society*, 4, 905–914. https://doi.org/10.1017/pds.2024.93
Dorst, K. (2011). The core of 'design thinking' and its application. *Design Studies*, 32 (6), 521–532. https://doi.org/10.1016/j.destud.2011.07.006
Gardenfors, P. (2004). Conceptual spaces: The geometry of thought. MIT Press.
Gero, J. S. (1990). Design prototypes: A knowledge representation schema for design. *AI Magazine*, 11 (4), 26–26.
Gero, J. S., & Kan, J. (2016). Empirical results from measuring design creativity: Use of an augmented coding scheme in protocol analysis. 1–8.
Goldschmidt, G. (2014). Linkography: Unfolding the design process. MIT Press.
Hay, L., Duffy, A. H., McTeague, C., Pidgeon, L. M., Vuletic, T., & Grealy, M. (2017). A systematic review of protocol studies on conceptual design cognition: Design as search and exploration. *Design Science*, 3, e10. https://doi.org/10.1017/dsj.2017.11
Kelly, N., & Gero, J. S. (2015). Situated interpretation in computational creativity. *Knowledge-Based Systems*, 80, 48–57. https://doi.org/10.1016/j.knosys.2014.12.005
Kelly, N., & Gero, J. S. (2022). Reviewing the concept of design frames towards a cognitive model. *Design Science*, 8, e30. https://doi.org/10.1017/dsj.2022.25
Kelly, N., Greentree, J., Sosa, R., & Evans, R. (2024). Automating useful representations of the design process from design protocols. In: Gero, J.S. (eds), *Design Computing and Cognition'24. DCC 2024*. Springer, pp. 3–20. https://doi.org/10.1007/978-3-031-71918-9_1
Kim, D. H., Genzola, R, R., Navea, M., Milovanovic, J., Gero, J. S., & Chrysikou, E. (2024). Learning preferences and individual differences in design fixation, Society for the Neuroscience of Creativity Conference 2023, San Francisco, CA, March 24. *Society for the Neuroscience of Creativity Conference 2023*, San Francisco, CA.
Salton, G. (1983). Introduction to modern information retrieval. McGraw Hill Book Co.
Schon, D. A. (1983). The Reflective Practitioner. *How Professionals Think in Action*. Basic Books.
Simon, H. A. (1969). The Sciences of the Artificial. MIT Press.
Simon, H. A. (1975). Style in design. In Eastman, C. (eds), *Spatial Synthesis in Computer-Aided Building Design*, 9, 287–309.
Valkenburg, R., & Dorst, K. (1998). The reflective practice of design teams. *Design Studies*, 19 (3), 249–271. https://doi.org/10.1016/S0142-694X(98)00011-8
Wasserman, S. & Faust, K. (1994). Social network analysis: Methods and applications. Cambridge University Press.

https://doi.org/10.1017/pds.2025.10301 Published online by Cambridge University Press