Shape Perception in Chemistry

Janna HASTINGS [a,b,c,1], Colin BATCHELOR [d] and Mitsuhiro OKADA [e]

[a] *Cheminformatics and Metabolism, European Bioinformatics Institute, Hinxton, UK*
[b] *Swiss Centre for Affective Sciences, University of Geneva, Switzerland*
[c] *Evolutionary Bioinformatics, Swiss Institute of Bioinformatics, Switzerland*
[d] *Royal Society of Chemistry, Cambridge, UK*
[e] *Department of Philosophy, Keio University, Tokyo, Japan*

Abstract. Organic chemists make extensive use of a diagrammatic language for designing, exchanging and analysing the features of chemicals. In this language, chemicals are represented on a flat (2D) plane following standard stylistic conventions. In the search for novel drugs and therapeutic agents, vast quantities of chemical data are generated and subjected to virtual screening procedures that harness algorithmic features and complex statistical models. However, in silico approaches do not yet compare to the abilities of experienced chemists in detecting more subtle features relevant for evaluating how likely a molecule is to be suitable to a given purpose. Our hypothesis is that one reason for this discrepancy is that human perceptual capabilities, particularly that of 'gestalt' shape perception, make additional information available to our reasoning processes that are not available to in silico processes. This contribution investigates this hypothesis.

Algorithmic and logic-based approaches to representation and automated reasoning with chemical structures are able to efficiently compute certain features, such as detecting presence of specific functional groups. To investigate the specific differences between human and machine capabilities, we focus here on those tasks and chemicals for which humans reliably outperform computers: the detection of the overall shape and parts with specific diagrammatic features, in molecules that are large and composed of relatively homogeneous part types with many cycles. We conduct a study in which we vary the diagrammatic representation from the canonical diagrammatic standard of the chemicals, and evaluate speed of human determination of chemical class. We find that human performance varies with the quality of the pictorial representation, rather than the size of the molecule. This can be contrasted with the fact that machine performance varies with the size of the molecule, and is of course impervious to the quality of diagrammatic representation.

This result has implications for the design of hybrid algorithms that take features of the overall diagrammatic aspects of the molecule as input into the feature detection and automated reasoning over chemical structure. It also has the potential to inform the design of interactive systems at the interface between human experts and machines.

Keywords. ontology, shape perception, cognition, spatial reasoning, logical reasoning, molecular graph

---

[1] Corresponding Author, e-mail: hastings@ebi.ac.uk

Introduction

> "A mind that has the ability to choose how it will represent a particular problem it needs to solve, choosing from a repertoire of representational capacities that include more analogical and more symbolic notations is more flexible, hence more 'intelligent' " [32]

Organic chemists make extensive use of chemical diagrams for designing, exchanging and analysing the features of chemicals. In this language, chemicals are represented on a flat (2D) plane following standard stylistic conventions [12]. The use of diagrammatic languages to concisely convey information for humans to process is an essential component of many sciences. In biology, pathway diagrams convey information about biological processes [14]. A good visualization of scientific information facilitates rapid understanding and can thereby lead to novel insights not otherwise possible [19]. One such example is Category Theory in mathematics, in which the use of diagrams is essential for representing mathematical properties and proofs [30].

In the search for novel drugs and therapeutic agents, large quantities of chemical data are generated. Interacting with these data and sifting a relevant subset (for a given problem) from the sizeable background is an ongoing challenge. Tools such as the *molecule cloud* can give an overview of a chemical dataset by showing common scaffolds sized for how often they appear in the dataset [9]. Many features of chemical entities have relevance on whether a given molecule is suited to a given purpose. Algorithmic and logic-based approaches are able to efficiently compute certain of these features, such as the presence of specific atoms or functional groups, overall mass and charge [18][2]. Algorithmic approaches can also gauge the overall shape of a molecule (at least in terms of delineating the outline of the three-dimensional space it fills) and calculate the mathematical similarity of that shape to that of other molecules or the reciprocity to potential binding sites [40,2,20]. Yet, many problems in chemical informatics remain difficult to efficiently automate over large molecular collections (e.g. finding maximal shared components between a set of molecular graphs [31], detecting all the cycles in a given molecule [3]).

In what follows, we focus on a class of problems that are known to be challenging for algorithmic solutions (in terms of efficiency), and yet are apparently straightforward for human chemists: detecting the overall *shape* and *class* of a presented molecule, in molecules that are large and composed of relatively homogeneous parts interconnected in cycles (such as the class of fullerene molecules [23]). As discussed in [18], determination of overall shape and chemical class for these classes of molecules is particularly challenging since the dense interconnection of the atoms in multiple fused cycles and the homogeneity of the atom environments. We hypothesise that a contributing factor in this performance discrepancy is that the use of a visual language in chemistry enables humans to directly harness the 'gestalt' or shape-detecting features of their visual perceptual machinery, *seeing the whole molecule at once* through the diagrammatic depiction, and therefore not needing to do the same sorts of computations that our algorithms need to do. If this line of thinking is correct, we should observe that the ability of humans to perform these tasks is affected by perturbances in the diagrammatic depiction more than in the size of the molecule. We conducted a study in which we time experi-

---

[2]As discussed in [18], we ignore statistical 'black box' approaches since they do not allow for explanations of their deductions and are not provably correct.

enced chemists performing a classification task on molecular diagrams with varied (a) diagrammatic faithfulness, and (b) size of the chemicals. We then evaluated the speed and accuracy of the chemists' performance given these variances.

The remainder of this document is organised as follows. In the next section, we give our experimental design in the context of some background information about chemical diagrams and the class of chemical problems that we will use as a case study. Thereafter, we present our results and discussion. With only three participants and only 30 diagrams included in our experiment, our results can be considered a pilot study rather than a conclusive investigation. However, we consider these preliminary findings suggestive of future research directions, and we go on to further speculate about the implications for the use of artificial intelligence in chemistry applications.

1. Methods

1.1. The diagrammatic language of chemistry

Molecular entities are commonly represented visually as connected graphs, in which the vertices represent atoms (or groups) and the edges chemical bonds [39]. Cheminformatics software use the underlying graph as chemical data structure that serves as input to algorithmic calculations of features of the chemicals. Logic-based approaches also use a graph-based underlying representation as input to automated reasoning processes [26]. For human consumption, however, the underlying graph is projected onto a two-dimensional plane for visual interpretation. This diagrammatic depiction is a core offering of almost all chemical databases, and professional chemists develop an aptitude at discerning molecular features via such representations. Some examples of chemical diagrams are illustrated in Figure 1.
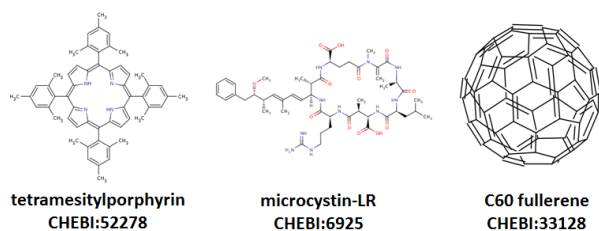

Figure 1. Some examples of chemical diagrams.

Chemical diagrams serve an invaluable purpose for chemists: they enable rapid evaluation of the overall chemistry of a given molecule, detection of errors or problems in the chemical structure being represented (e.g. infeasibility or chemical instability), and assessment of the properties or classifications that are relevant for the given molecule.

Chemical diagrams, like maps, represent *spatial* information. We have earlier referred to such spatial representations such as street maps, chemical diagrams, and engineering design models as *structural diagrams* [12], and they were called *analogical representations* in [36]. Here, we will focus not on the structural associations that we

highlighted before, but on the features of the overall shape and layout of chemicals that are available in chemical diagrams.

Molecular flexibility is also very important for molecular shape [13], as one 2D depiction can, through flexibility, yield many different 3D conformers that have vastly different properties in vitro. Such flexibility is not explicitly represented in 2D illustrations of molecules, but can be inferred from such representations given appropriate chemical knowledge.

1.2. Chemical shape perception task

For our experiment, we have deliberately chosen classes of molecule that are known to be challenging to represent with logic-based automated reasoning approaches. Earlier, we have conducted an evaluation of the capabilities of algorithmic and logic-based approaches to reasoning tasks with molecular structures in [18]. The classes we selected for use in this task are:

1. Macrocyclic molecules, including calixarenes;
2. Polycyclic cages, including several differently sized fullerenes;
3. Shape-characterised molecules such as the catenanes and molecular knots;
4. Molecules that were not members of the above three classes as 'controls'.

Macrocyclic molecules are molecules that form a large cyclic structure composed of linkages of smaller functional groups. Polycyclic cages are molecules that are composed entirely of cycles that are fused together in such a way as to form an overall cage-like structure, which is a feature that has interesting applications in medicinal chemistry and in materials science as the structure can serve to protect or capture a smaller molecule on the inside, or be engineered to lengthy tubes that are very strong. Examples are the fullerenes, cucurbiturils (named for their similarity to pumpkins), nanotubes, and small regular compounds such as cubane. Such nanomaterials have recently shown promise in the challenge of capturing highly volatile nerve agents and thereby preventing damage in vivo [22]. Molecules with specific shapes are of interest in the development of molecular machines, including the presence of stationary and movable parts, and the ability to respond with controlled movements to the external environment. Molecules that are mechanically interlocked—such as bistable rotaxanes and catenanes are some of the most intriguing systems in this area because of their capacity to respond to stimuli with controlled mechanical movements of one part of the molecule (e.g. one interlocked ring component) with respect to the other stationary part [10]. Similarly, molecules which display unusual energetic properties by virtue of their overall shape, such as molecular Möbius strips and trefoil knots, are an active research area for many novel applications, and in many cases mimic the extraordinary properties of biomolecular machinery such as active sites within protein complexes [34,42].

We selected five individual molecule types for the first two classes (macrocyclic molecules and cages). For the shape-characterised molecules, we were not able to find as many representatives in public chemical databases (our main source was the ChEBI database [15]), therefore we selected only four examples. Eight molecules that were not members of any of the three target classes but which were highly similar to one of the selected molecules (based on cheminformatics similarity scoring using Tanimoto over the molecular fingerprint, as implemented in OrChem [33]). Molecules were selected ranging from small to large, as measured in terms of counts of non-hydrogen atoms.

A randomly selected subset of eight of those 22 molecules was then subjected to diagrammatic distortion. Different distortion mechanisms were used. Firstly, the original molecule was computationally assigned a 3D conformation, which was then projected back onto a 2D diagram (a common outcome of computational processing of chemicals originally drawn by human chemists). Secondly, computational procedures for 'clean' 2D diagram generation were used. Finally, some of the diagrams were subjected to image processing to obscure the standard chemical representation either through blurring or shape-based transformation. The total number of diagrams was thus 30. The full set of molecules is shown in Figure 2.
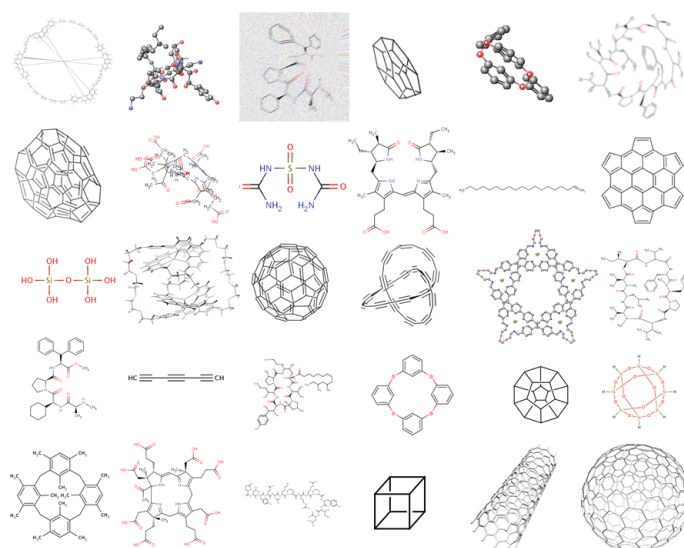

Figure 2. The molecule diagrams used in the experiment, including those showing distortions.

These diagrams were then displayed to the three experienced chemist participants in a random sequence[3]. For each diagram, the chemist was asked to determine the chemical class of the molecule, presented with the three classes, a fourth option 'none of the above,' and a final option 'unable to tell from this diagram.' Participants were timed as they completed the task, and their accuracy and agreement were calculated. Figure 3 shows a screenshot of the interface we developed in order to complete the perceptual task.

---

[3]There were three participants, each of whom had an academic background in chemistry and interacted with chemical data on a daily basis. The participants were explained the purpose of the experiment and each gave their informed consent. All data were stored anonymously and securely.
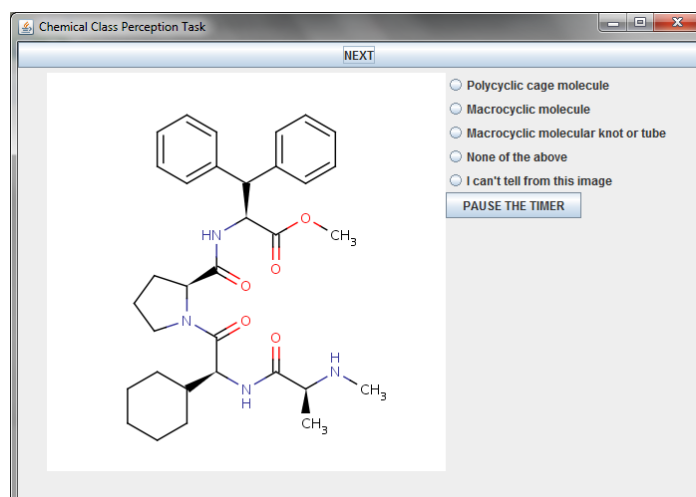
Figure 3. The chemical perception task interface, showing the chemical diagram and class selection options.

2. Results

2.1. The effect of image distortion on performance

Image distortion had a significant effect on the accuracy of the chemical raters in choosing the correct classification for the classes. Figure 4 (a)[4] shows a boxplot of the classification task accuracy for the standard images as compared to the distorted images. The time taken (Figure 4 (b)) shows less of an effect than the accuracy, with the means not significantly different but the variance much larger in the case of the distorted images.

Ordinarily, chemists would look for additional information in case they encounter a partially obscured image and needed to determine the chemical class. Therefore, we do not restrict here our measure of accuracy to the percentage of correct classifications. Agreement between chemists in a classification task is an alternative measure of accuracy, which is especially useful in case the correct classification is not known in advance, but can supplement the known accuracy score used above with a clue as to the difficulty of the task. It might, for example, have been the case that the chemists had all agreed on incorrect classifications for the distorted images, leading to low accuracy but high agreement. However, agreement also differed strongly between the non-distorted and the distorted set of images, with the distorted images having a much lower agreement as measured by Cohen's Kappa statistic for multiple raters [6]. For the non-distorted pictures, the kappa was 0.88. For the distorted pictures, the kappa was 0.46%.

2.1.1. The effect of size on performance

The scatter plot in Figure 5 (a) shows that size did not have a large impact on the time taken to perform the task. The red correlation line shows a weak positive correlation

---

[4]For space considerations, we do not present the full raw data result table here. However, this is available on request.
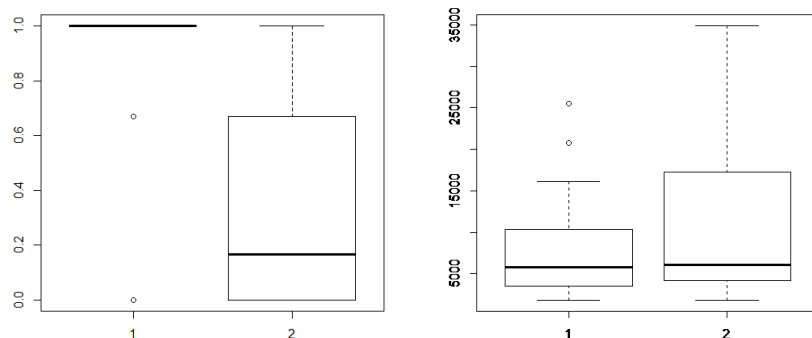
Figure 4. The display shows boxplots of (a) accuracy and (b) time taken (in ms), comparing the results for good vs. obscured visual layouts of the chemicals
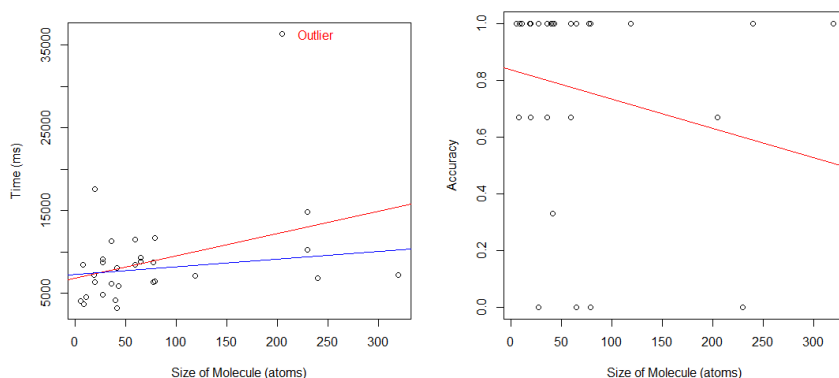

Figure 5. Scatter plot of average (a) time to complete task (ms), and (b) accuracy, against size of the molecule.

between size and time taken to complete the task. However, this correlation is largely influenced by one data point, which itself depends on just one data point. The blue line shows the much weaker correlation that results from excluding the single outlier from the analysis.

Figure 5 (b) shows that accuracy was slightly anti-correlated with the size of the molecule, but this effect is not significant, with the p-value of the correlation only 0.24, and the 95% confidence interval for the correlation coefficient was from -0.54 to 0.15.

These results can be compared to algorithmic approaches and logic-based approaches for the relevant sort of feature detection that would be required to automatically compute the same task, i.e. automatic classification into the correct class based on chemical structures. Unfortunately, there is not yet any available generic system that is

able to perform the classifications tasks that were used in this experiment with which we may have compared the performance to our human experts. Indeed, as discussed in [18] our research has the long-term objective of enabling the development of just such a system, however, at this preliminary stage we do not have an available benchmark but must instead look to the performance profiles of algorithms that are known to be relevant.

2.2. Algorithmic cheminformatics approaches

The relevant algorithms that would be required to detect the classes specified include the detection of *subgraph isomorphism* and finding the *smallest set of rings* [18,43]. These algorithms are known to scale supralinearly in the number of atoms. For example, subgraph isomorphism in the general case is known to be NP-complete [7], although optimisations exist for various sub-classes of molecules, such as those that are planar [8].

For the particularly shape-defined classes, shape similarity algorithms on molecular structures exist that use ray-tracing of the projected surfaces of molecules to estimate the overall shape of the molecule and use that as a descriptor e.g. in virtual screening [2]. These methods depend on 3D conformer though, and for flexible molecules many conformers may result from the same 2D diagrammatic depiction, dramatically decreasing the performance of the algorithm.

Furthermore, a separate algorithm implementing a check on the rules of class membership would need to be hand-written for each of the three class types used in this task (macrocyclic, cage, shape-defined). This hampers the extensibility and flexibility of a system that needs to classify molecules in the general case [18,25]. On the other hand, logic-based systems address these objectives of being generic, extensible and flexible.

2.3. Logic-based approaches

The popular Web Ontology Language, OWL [11], is highly efficient in representing tree-like structures, but is unable to correctly represent cyclic structures [16]. A first-order logic programming based formalism has been proposed specifically for the case of representing chemical structures [26,25]. These description graph logic programs (DGLP) are able to represent objects whose parts are interconnected in arbitrary ways, including cyclic structures. The decidability of logic programs do not rely on the tree-model property that underlies the description logics behind OWL. However, representation of classes with more advanced overall topological features such as polycyclic cages is beyond the expressivity of DGLP as it requires quantification over all atoms in a molecule rather than specific atoms, parts or properties within the molecule.

Perhaps motivated by similar concerns on the limits of the logic-based approaches underlying languages such as OWL, Maojo *et al.* propose a 'morphospatial' approach to ontology with application in the nanomaterials domain [27]. Shape features are explicitly encoded in their ontology alongside other features such as composition. However, this approach merely pushes the problem onto those computational methods that are needed to derive the shape features automatically from some representation of the input chemical structure and thereby assign appropriate ontological categories to nanomolecular structures.

An approach for the representation of the overall structure or topology of highly symmetrical polycyclic molecules is described in [17,23]. There, the authors propose us-

ing a combination of monadic second-order logic and ordinary OWL, with a heterogeneous logical connection framework used to bridge between the two formalisms. This approach has not yet been implemented in practice, but shows promise for logical reasoning over features involving regularity in the overall structure of molecules. However, arbitrary entailment in monadic second-order logic is known to be computationally expensive[5].

Spatial logics and spatial axiomatizations have been advanced in which it is possible to perform computational deductive reasoning [5,41]. However, it is not immediately straightforward to represent the problem of determining from an arbitrary chemical graph whether it is a member of the class of fullerenes (for example) as a spatial reasoning problem. We will develop this research question further in future work.

3. Discussion and Conclusions

While this study is small and exploratory in nature only, our results provide tentative support for a role for perception in human performance in the presented classification decision task, in that observed performance appeared to decrease with the quality of the diagrammatic representation rather than the size of the molecule. On the other hand, it is known that the best algorithmic and logical approaches to solving these particular tasks scale dramatically in the size of the molecule, rendering their habitual application to large numbers of molecules in a database problematic.

Larkin and Simon [24] attribute observed efficiencies of diagrammatic reasoning relative to non-diagrammatic reasoning to efficiencies in searching and inference in the reprentation space compared to that of a non-diagrammatic representation space, e.g. axioms. This may indeed be the root explanation, but it doesn't give guidance on how best to expose the representational efficiency that humans have (the ability to perceive the overall shape and connectivity in molecular diagrams) to computational processes. Traditional logical reasoning relies on linguistic or symbolic representation of the properties of objects together with the rules for deriving inferences on those properties. By contrast, diagrammatic representation can explicitly encode the relevant properties of objects and their background constraints such that the needed inferences can be directly drawn from the spatial constraints evident in the illustration [32], known as the "free ride" property.

Systems have been developed that enable the representation of logical axioms diagrammatically and the formalisation of accompanying reasoning systems to the extent that diagrammatic and traditional syllogistic reasoning can be combined in order to serve as an aid for human capability [29,28]. Such logical diagrammatic representations do not correspond directly to portions of reality, as the diagrammatic representations of chemicals correspond to classes of chemicals, but the correspondence is still analogical, i.e. by analogy. For example, Euler diagrams represent axioms such as *All A are B* as a smaller circle *A* entirely enclosed in a larger circle *B* [35]. This is analogous to *spatial inclusion*, as (for example) a smaller fullerene molecule can be fully enclosed in a larger fullerene molecule [23], and we could make corresponding statements such as *All atoms in molecule A are IN$_{SPATIALLY}$ molecule B*.

---

[5]Automated theorem provers such as LEO-II (http://www.ags.uni-sb.de/~leo/) are able to approximate some aspects of entailment checking.

Where perception is used as an aid to reasoning, care must be taken in the choice of the visual representation. For example, when visual diagrams are used as an aid to human logical reasoning, it has been found that Euler diagrams are more effective than Venn diagrams [28]. Irrelevant and distracting visual detail acts as a hindrance to reasoning rather than as an aid [21]. In the chemistry domain, for the class of classification problems we are interested in, this may be particularly important. Exposing the specifically visual information of a chemical diagram to computational processes would introduce additional constraints on the representation of the chemicals that currently only obtain in case the representation is intended for human consumption. Visual inference can sometimes be much more expensive than normal inference in the corresonding axiomatization, especially when the visual information is incomplete or, as we have tested, perturbed [1]. Adherence to standards for clear and unambiguous diagrammatic representation such as those put forward in [4] would go some way to address this concern in the chemistry domain.

In chemical similarity searching and bioactivity predictive modelling, quantitative shape-based 3D descriptors have met with mixed results stemming, on the one hand, from their greater computational cost than their 2D counterparts, and on the other hand, from the additional 'noise' that they can introduce in flexible molecules due to the variety of conformations [40]. One direction for our future research will be to evaluate the performance of these shape-based descriptors in assigning shape classes to molecules, such as 'spherical' and 'cubic'. We are not aware of any existing work that applies this type of descriptor to the problem of structure-based chemical classification.

Our result emphasises the need for hybrid reasoning systems in chemistry that are able to combine features derived diagrammatically from visual representations of the molecule with the now-standard logic-based and algorithmic reasoning over the graph-based structure. Such hybrid systems have been advanced in other domains. For example, the Vivid system offers some diagrammatic reasoning capability alongside logical reasoning capability [1]. However, this system depends on algorithmic processes that "observe" pre-defined features in the diagrams included in the system capability. In the case of the chemical diagrams that form our case study here, some features are features of the *whole* diagram for which computational "observation" algorithms do not (to the best of our knowledge) yet exist. Research in machine vision may yield some methods that can be harnessed in pursuit of this objective [38].

Sloman [37] speculates that the ability to integratively process different types of representation with correspondingly different reasoning tasks might be a distinctive feature of intelligence in general; it is certainly a feature of human intelligence.

References

[1] Arkoudas, K., Bringsjord, S.: Vivid: A framework for heterogeneous problem solving. Artificial Intelligence (Jun 2009), `http://dx.doi.org/10.1016/j.artint.2009.06.002`

[2] Ballester, P.J., Richards, W.G.: Ultrafast shape recognition for similarity search in molecular databases. Proc. R. Soc. A 463, 1307–1321 (2007)

[3] Berger, F., Flamm, C., Gleiss, P.M., Leydold, J., Stadler, P.F.: Counterexamples in chemical ring perception. J Chem Inf Comput Sci 44, 323–331 (2004)

[4] Brecher, J., Degtyarenko, K.N., Gottlieb, H., Hartshorn, R.M., Hellwich, K.H., Kahovec, J., Moss, G.P., McNaught, A., Nyitrai, J., Powell, W., Smith, A., Taylor, K., Williams, A., Yerin, A., Town, W.: Graphical representation standards for chemical structure diagrams (IUPAC recommendations 2008). Pure and Applied Chemistry 80, 277–410 (2008)

[5] Cabedo, L.M., Escrig, M.T.: A qualitative theory for shape representation and matching for design. In: R. Lopez de Mantaras and L. Saitta, editors, Proceedings of the 16th European Conference on Artificial Intelligence (2004)

[6] Conger, A.: Integration and generalisation of kappas for multiple raters. Psychological Bulletin 88, 322–328 (1980)

[7] Cook, S.A.: The complexity of theorem-proving procedures. In: Proc. 3rd ACM Symposium on Theory of Computing. p. 151158 (1971)

[8] Eppstein, D.: Subgraph isomorphism in planar graphs and related problems. Journal of Graph Algorithms and Applications 3, 1–27 (1999)

[9] Ertl, P., Rohde, B.: The molecule cloud - compact visualization of large collections of molecules. Journal of Cheminformatics 4(1), 12 (2012), `http://www.jcheminf.com/content/4/1/12`

[10] Forgan, R.S., Sauvage, J., Stoddart, J.F.: Chemical topology: Complex molecular knots, links, and entanglements. Chemical Reviews 111, 5434–5464 (2011)

[11] Grau, B.C., Horrocks, I., Motik, B., Parsia, B., Patel-Schneider, P., Sattler, U.: OWL 2: The next step for OWL. Web Semantics 6, 309–322 (November 2008), `http://portal.acm.org/citation.cfm?id=1464505.1464604`

[12] Hastings, J., Batchelor, C., Neuhaus, F., Steinbeck, C.: What's in an 'is about' link? Chemical diagrams and the IAO. In: Proceedings of the International Conference on Biomedical Ontology (ICBO2011), Buffalo, USA (2011)

[13] Hastings, J., Batchelor, C., Schulz, S.: Parts and wholes, shapes and holes in living beings. In: Proceedings of the SHAPES 1.0 workshop, Karlsruhe, Germany. CEUR-WS volume 812. (2011)

[14] Hastings, J., Batchelor, C., Schulz, S., Jansen, L.: Collective bio-molecular processes: The hidden ontology of systems biology. Proceedings of the UMoCoP workshop, July 2012, Birmingham UK (2012)

[15] Hastings, J., de Matos, P., Dekker, A., Ennis, M., Harsha, B., Kale, N., Muthukrishnan, V., Owen, G., Turner, S., Williams, M., Steinbeck, C.: The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. Nucleic Acids Research Database issue, gks1146 (2013)

[16] Hastings, J., Dumontier, M., Hull, D., Horridge, M., Steinbeck, C., Sattler, U., Stevens, R., Hörne, T., Britz, K.: Representing chemicals using OWL, description graphs and rules. In: Proc. of OWL: Experiences and Directions (OWLED 2010) (2010)

[17] Hastings, J., Kutz, O., Mossakowski, T.: How to model the shapes of molecules? combining topology and ontology using heterogeneous specifications. In: Proceedings of the DKR Challenge Workshop, Banff, Alberta, Canada, June 2011. (2011)

[18] Hastings, J., Magka, D., Batchelor, C., Duan, L., Stevens, R., Ennis, M., Steinbeck, C.: Structure-based classification and ontology in chemistry. Journal of Cheminformatics 4(1), 8 (2012), `http://www.jcheminf.com/content/4/1/8`

[19] Hey, A.J.G., Tansley, D.S.W., Tolle, K.M.: The Fourth Paradigm: Data-Intensive Scientific Discovery. Microsoft Research, Redmond, WA, USA (2009)

[20] Kim, S., Bolton, E., Bryant, S.: PubChem3D: Shape compatibility filtering using molecular shape quadrupoles. Journal of Cheminformatics 3(1), 25 (2011), `http://www.jcheminf.com/content/3/1/25`

[21] Knauff, M., Johnson-Laird, P.N.: Visual imagery can impede reasoning. Memory and Cognition 30, 363–371 (2002)

[22] Kowalczyk, P., Gauden, P.A., Terzyk, A.P., Neimark, A.V.: Screening of carbonaceous nanoporous materials for capture of nerve agents. Phys. Chem. Chem. Phys. pp. – (2013), `http://dx.doi.org/10.1039/C2CP43366D`

[23] Kutz, O., Hastings, J., Mossakowski, T.: Modelling Highly Symmetrical Molecules: Linking Ontologies and Graphs Artificial Intelligence: Methodology, Systems, and Applications. Lecture Notes in Computer Science, vol. 7557, chap. 11, pp. 103–111. Springer Berlin / Heidelberg, Berlin, Heidelberg (2012),

http://dx.doi.org/10.1007/978-3-642-33185-5_11

[24] Larkin, J., Simon, H.: Why a diagram is (sometimes) worth ten thousand words. Cognitive Science 11, 65–99 (1987)

[25] Magka, D.: Ontology-based classification of molecules: A logic programming approach. In: Proceedings of the SWAT4LS conference, 30 November 2012, Paris, France. p. . (2012)

[26] Magka, D., Motik, B., Horrocks, I.: Modelling structured domains using description graphs and logic programming. Tech. rep., Department of Computer Science, University of Oxford (2011)

[27] Maojo, V., Fritts, M., Martín-Sánchez, F., de la Iglesia, D., Cachau, R.E., García-Remesal, M., Crespo, J., Mitchell, J.A., Anguita, A., Baker, N., Barreiro, J.M., Benitez, S.E., de la Calle, G., Facelli, J.C., Ghazal, P., Geissbühler, A., Gonzalez-Nilo, F., Graf, N.M., Grangeat, P., Hermosilla, I., Hussein, R., Kern, J., Koch, S., Legré, Y., López-Alonso, V., López-Campos, G., Milanesi, L., Moustakis, V., Munteanu, C.R., Otero, P., Pazos, A., Pérez-Rey, D., Potamias, G., Sanz, F., Kulikowski, C.A.: Nanoinformatics: developing new computing applications for nanomedicine. Computing 94(6), 521–539 (2012)

[28] Mineshima, K., Okada, M., Takemura, R.: Two types of diagrammatic inference systems: Natural deduction style and resolution style. In: Diagrams 2010, Lecture Notes In Artificial Intelligence 6170. Springer (2010)

[29] Mineshima, K., Okada, M., Takemura, R.: A generalized syllogistic inference system based on inclusion and exclusion relations. Studia Logica (2012)

[30] P, V., J, F., N, R.: Understanding visualization: A formal approach using category theory and semiotics. IEEE Trans Vis Comput Graph. Sep 21. (2012)

[31] Raymond, J.W., Willett, P.: Maximum common subgraph isomorphism algorithms for the matching of chemical structures. Journal of Computer-Aided Molecular Design 16, 521–533 (2002)

[32] Recanati, C.: Hybrid reasoning and the future of iconic representations. In: Proceedings of the 2008 conference on Artificial General Intelligence 2008: Proceedings of the First AGI Conference. pp. 299–310. IOS Press, Amsterdam, The Netherlands, The Netherlands (2008), http://dl.acm.org/citation.cfm?id=1566174.1566202

[33] Rijnbeek, M., Steinbeck, C.: OrChem - An open source chemistry search engine for Oracle(R). Journal of Cheminformatics 1(1), 17 (2009), http://dx.doi.org/10.1186/1758-2946-1-17

[34] Rzepa, H.S.: Wormholes in chemical space connecting torus knot and torus link π-electron density topologies. Phys. Chem. Chem. Phys. pp. 1340–1345 (2009)

[35] Sato, Y., Mineshima, K., Takemura, R., M.Okada: On the cognitive efficacy of euler diagrams in syllogistic reasoning: A relational perspective. In: Euler Diagram Workshop 2012, CEUR 854 (2012)

[36] Sloman, A.: Interactions between philosophy and ai: The role of intuition and non-logical reasoning in intelligence. In: Proceedings of the Second International Joint Conference on Artificial Intelligence. p. . (1971)

[37] Sloman, A.: Musings on the roles of logical and non-logical representations in intelligence. In: Glasgow, J., Narayanan, H., Chandrasekaran (eds.) Diagrammatic Reasoning: Computational and Cognitive Perspectives, p. . AAAI Press (1995)

[38] Sun, M., Bradski, G., Xu, B.X., Savarese, S.: Depth-encoded hough voting for coherent object detection, pose estimation, and shape recovery. In: ECCV (2010)

[39] Trinajstic, N.: Chemical graph theory. CRC Press, Florida, USA (1992)

[40] Venkatraman, V., Chakravarthy, P., Kihara, D.: Application of 3d zernike descriptors to shape-based ligand similarity searching. Journal of Cheminformatics 1(1), 19 (2009), http://www.jcheminf.com/content/1/1/19

[41] Vincent Dugat, P.G., Larvor, Y.: Qualitative geometry for shape recognition. Applied Intelligence 17, 253–263 (2002)

[42] Wannere, C.S., Rzepa, H.S., Rinderspacher, B.C., Paul, A., III, H.F.S., v. R. Schleyer, P., Allan, C.S.M.: The geometry and electronic topology of higher-order Möbius charged annulenes. J. Phys. Chem. 113, 11619–11629 (2009)

[43] Wegner, J.K., Sterling, A., Guha, R., Bender, A., Faulon, J.L., Hastings, J., O'Boyle, N., Overington, J., Van Vlijmen, H., Willighagen, E.: Cheminformatics. Communications of the ACM 55(11), 65–75 (2012)