

CATALYSIS

Catalyst: The Metaphysics of Chemical Reactivity

Leroy Cronin,^{1,*}
S. Hessam M. Mehr,¹
and Jarosław M. Granda¹

Leroy (Lee) Cronin has been thinking about programming chemistry since he was 8 years old, when he got his first computer and chemistry set. He is now the Regius Professor of Chemistry at the University of Glasgow. He and his team are trying to make artificial life forms, find alien life, explore the digitization of chemistry, and construct chemical computers.

S. Hessam M. Mehr is a post-doctoral researcher at the University of Glasgow. He earned his PhD at the University of British Columbia (2017). His research focuses on probabilistic and information-theoretic methods in chemistry and the discovery of chemical novelty.

Jarosław M. Granda is a postdoctoral research associate at the University of Glasgow. He obtained his PhD at the Institute of Organic Chemistry of the Polish Academy of Sciences (2014). His research interests focus on the automation of organic synthesis and artificial-intelligence-driven systems for chemical discovery.

How do we know where to look to discover new types of reactions, molecules, and reactivity in chemistry? Chemistry is a practical subject, and most of our knowledge originates from laboratory experiments. The fundamental building blocks of chemistry, described by the periodic table, allow us to draw general conclusions regarding the reactivity of a given compound as a function of its elemental composition.¹ However, the notion of

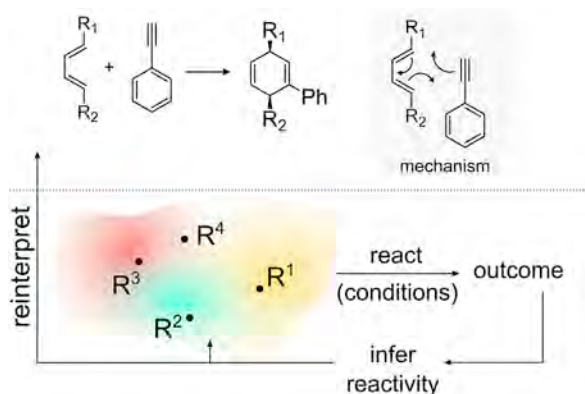
reactivity has been refined over the years within this framework and is supported by theories of bonding and chemical structure. But how can we go beyond our current knowledge? Is it possible to develop an abstraction of chemistry beyond the current rules of reactivity—a new meta-chemical reality? Could we use this new abstraction to discover entirely new knowledge that can be translated back to chemical space so that chemists can then apply it to build new molecules and engineer new reactions? We would suggest that we might need to develop the metaphysics of chemistry before jumping on the big-data and artificial-intelligence bandwagon.

Metaphysics is a branch of philosophy concerned with the nature of existence and reality, according to which the application of logic alone should allow the construction of the reality around you in a virtual metaphysical box giving the laws of the universe, including chemistry. This approach is important because it accepts that there is a reality (causation), and it forces the observer or experimentalist to strip away assumptions of cause and effect as opposed to real observations. In many sciences (chemistry is not immune), the assumption of correlation and causation corrupts what knowledge we have and can be a source of bias.² This bias has been incorporated into our models and theories over time and thus distracts the observer from challenging the models because we are told that things won't work outside the walled garden. Typically, these assumptions go unchallenged until a mistake is made—a chance event or serendipitous discovery—or the basis for the model is found to be incorrect.

The current optimism and excitement surrounding the hope that developments in algorithm development, machine learning, and artificial intelli-

gence will have a profound impact on the world are also affecting chemical and materials science.³ From exploring databases to developing rule-based retrosynthesis,⁴ the application of these approaches to chemistry is promising, but it is not without problems.⁵ First, finding the correct way to represent chemical knowledge is hard.⁶ Most approaches use structural representations or an encoding of these representations, such as SMILES or InChI. Second, chemical space (we define this here as being some function of molecular weight and elemental composition) is very large and sparse, and although vast areas contain only a few molecules, some parts are highly populated with molecules. Third, chemistry costs both time and physical resources, resulting in a limit on the number of experiments that can be done. Fourth and finally, many of the chemical databases that hold reaction data are incomplete, and much of the data are not validated. In some cases, the data are erroneous, and many of these databases are not freely available. Perhaps most frustratingly, only positive results are reported. This is an important problem given that many machine-learning techniques, e.g., neural networks, require a very large amount of data for training, and the quality of the resulting model critically depends upon the quality of the input data.⁷ So how can we develop insights when our data are limited and the quality is variable? To address the issue of poor-quality data, we have been developing closed-loop approaches with autonomous robots that are able to decide which experiments to do, collect the data, and then make a decision about what experiment to do next.^{8,9} These are likely to be much more efficient than current approaches because they will improve as the experiments are done, thereby reducing the amount of resources required. This is because the budget of the total number of reactions can be better utilized on





Scheme 1. Exploring Chemical Space

This scheme shows well-defined chemical language (top) and the reactions of molecules, and the resulting mechanism can be expressed with process variables (bottom) to react the reagents together, detect the outcome, infer reactivity, and translate to chemical knowledge.

the reactions that are more likely to match the search goal. With this in mind, instance-based algorithms, such as simulated annealing or genetic algorithms, can be used for choosing the next experiments with data obtained from previous experiments. These approaches rely on successive experiments aiming for improvement in the overall trend by selecting each subsequent experiment in a statistical approach.

Model-based approaches also use feedback from previous experiments with the goal of improving the outcome of the ongoing search, but the data are also used for constructing a model of the world. Such models are dependent upon the previous data.

Linking experimental generation and model development can improve knowledge of the space and dramatically increase the insights gained as a function of the experiments. This is analogous to how a good human experimenter will work, and even though humans are normally superior because of their depth of knowledge, they also can be limited by assumptions or bias about what experiments should not be done.¹⁰ This means that a well-designed exploration system can challenge the assumptions built into the

model and update them. Therefore, human experimenters, if they are willing, can also update their own bias and knowledge with regard to an agnostic model.

Discovery in chemistry falls mainly into one of four areas: new molecules, new reactions, new reactivity, and finally new physical properties of the resulting compounds or materials. Establishing new reactivity leads to new reactions, which also leads to new molecules. This is therefore the order of impact for discoveries in terms of the amount of chemical knowledge that they contribute. Such findings must, by definition, belong outside the known or predictable; they are outliers and as such can oppose conventions, assumptions, and biases. We can use the idea of an outlier to define a novel discovery in practical terms as any information about the chemical space that exhibits sufficiently different outcomes from prediction. By taking this approach, we can reformulate the problem of chemical discovery. In this way, we set up a series of experiments, predict the outcomes with current knowledge, and flag those instances that diverge from the prediction. Of course, novelty in chemistry has been hard to argue for because the assertion of novelty is only as good as the knowledge held

by the person or expert system that asserts that the observation is novel. This is only a practical and not a theoretical problem. We define an observation as novel if it cannot be predicted with prior knowledge. In chemistry, we have to add practical filters so we don't confuse novelty with issues of reaction sensitivity to initial conditions, issues with reagent quality, fine control over reaction processing, and so on.

A significant change to discovery in chemistry could be made possible if automation and/or "smart" algorithms could help with the design of experiments to focus on the areas of chemical space that have been poorly explored, found to be unreliable, or aiming for previously inaccessible areas (reactions that are too sensitive to initial conditions or give stochastic outcomes). The key issue is that we don't currently report failed experiments, yet ironically, mapping failure is crucial in increasing the chance of discovery. This means that there is a massive opportunity for chemists to not only record failure but also find a way to uniformly present it in a standard way to aid in the design of experiments and the input experiment list. In addition, the development of new material-handling systems and sensor arrays will be crucial so that the robotic system doing the chemistry could be more efficiently directed.

By developing the meta-physics of chemistry and chemical reactivity, we should be able to establish a new set of chemical ontologies that relate back to the practical core operations but that also can be translated into molecular structures and the discovery of function. The truth of chemistry lies with finding the intrinsic reactivity of the input chemicals and then encouraging or enabling reactivity by process control (Scheme 1). This could be controlling temperature, adding a catalyst, changing solvent, and so on. Although the new discovery and reaction should

be translatable to chemical bonding theory, chemists need to grapple with the fact that the application of the current rules will not allow discovery; instead, it will act to restrict rules to those that are known. So, chemical discovery requires that the current rules be updated or broken or that new ones be made. The discovery of Diels-Alder or cross-coupling reactions is an excellent example of a new rule that was discovered without any prior warning.

The conclusion is that without a deeper development of a meta-physics of chemistry, the use of big data and artificial intelligence will just tell us what we already know we know, and maybe predictable extensions, rather than enable discovery. The challenge for the chemist is not the use of artificial intelligence but the intelligent use of algorithms and automation for novel discoveries rather than just new molecules that are predictable. This development is crucial if chemical technologies are to shake the perceived failure of the combinatorial synthesis revolution. Ultimately, the development of such

tools should build on the creativity of the chemist and allow discoveries and developments that would not have been possible in isolation. With such approaches, the chemist will be able to boldly go into the unknown and actively seek chemical novelty.

1. Scerri, E. (2005). Some aspects of the metaphysics of chemistry and the nature of the elements. *PhilSci*, <http://philsci-archive.pitt.edu/2213/>.
2. Waldmann, M.R. (2007). Combining versus analyzing multiple causes: how domain assumptions and task context affect integration rules. *Cogn. Sci.* 31, 233–256.
3. Tabor, D.P., Roch, L.M., Saikin, S.K., Kreisbeck, C., Sheberla, D., Montoya, J.H., Dwaraknath, S., Aykol, M., Ortiz, C., Tribukait, H., et al. (2018). Accelerating the discovery of materials for clean energy in the era of smart automation. *Nat. Rev. Mat.* 3, 5–20.
4. Klucznik, T., Mikulak-Klucznik, B., McCormack, M.P., Lima, H., Szymkuć, S., Bhowmick, M., Molga, K., Zhou, Y., Rickershauser, L., Gajewska, E.P., et al. (2018). Efficient syntheses of diverse, medically relevant targets planned by computer and executed in the laboratory. *Chem* 4, 522–532.
5. Gómez-Bombarelli, R., Wei, J.N., Duvenaud, D., Hernández-Lobato, J.M., Sánchez-Lengeling, B., Sheberla, D., Aguilera-Iparraguirre, J., Hirzel, T.D., Adams, R.P., and Aspuru-Guzik, A. (2017). Automatic chemical design using a data-driven continuous representation of molecules. *arXiv*. <https://arxiv.org/abs/1610.02415>.
6. Segler, M.H.S., Preuss, M., and Waller, M.P. (2018). Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* 555, 604–610.
7. Granda, J., Donina, L., Dragone, V., Long, D.-L., and Cronin, L. (2018). Controlling an organic synthesis robot with machine learning to search for new reactivity. *Nature* 559, 377–381.
8. Points, L.J., Taylor, J.W., Grizou, J., Donkers, K., and Cronin, L. (2018). Artificial intelligence exploration of unstable protocells leads to predictable properties and discovery of collective behavior. *Proc. Natl. Acad. Sci. USA* 115, 885–890.
9. Parrilla-Gutierrez, J.M., Tsuda, S., Grizou, J., Taylor, J., Henson, A., and Cronin, L. (2017). Adaptive artificial evolution of droplet protocells in a 3D-printed fluidic chemorobotic platform with configurable environments. *Nat. Commun.* 8, 1144.
10. Duros, V., Grizou, J., Xuan, W., Hosni, Z., Long, D.-L., Miras, H.N., and Cronin, L. (2017). Human versus Robots in the Discovery and Crystallization of Gigantic Polyoxometalates. *Angew. Chem. Int. Ed.* 56, 10815–10820.

¹WestCHEM, School of Chemistry, University of Glasgow, Glasgow G12 8QQ, UK

*Correspondence: lee.cronin@glasgow.ac.uk
<https://doi.org/10.1016/j.chempr.2018.07.008>