This is an open access article published under a Creative Commons Attribution (CC-BY) License, which permits unrestricted use, distribution and reproduction in any medium, provided the author and source are cited.

JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

Cite This: J. Chem. Inf. Model. 2019, 59, 2664−2671

pubs.acs.org/jcim

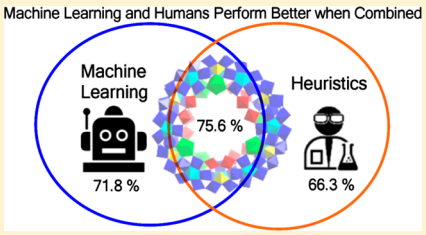
Intuition-Enabled Machine Learning Beats the Competition When Joint Human-Robot Teams Perform Inorganic Chemical Experiments

Vasilios Duros, Jonathan Grizou, Abhishek Sharma, S. Hessam M. Mehr, Andrius Bubliauskas, Przemysław Frei, Haralampos N. Miras, and Leroy Cronin*

School of Chemistry, University of Glasgow, University Avenue, Glasgow G12 8QQ, United Kingdom

Supporting Information

ABSTRACT: Traditionally, chemists have relied on years of training and accumulated experience in order to discover new molecules. But the space of possible molecules is so vast that only a limited exploration with the traditional methods can be ever possible. This means that many opportunities for the discovery of interesting phenomena have been missed, and in addition, the inherent variability of these phenomena can make them difficult to control and understand. The current state-of-the-art is moving toward the development of automated and eventually fully autonomous systems coupled with in-line analytics and decision-making algorithms. Yet even these, despite the substantial progress achieved recently, still cannot easily tackle large combinatorial spaces, as they are limited by the lack of high-quality data. Herein, we explore the utility of active learning methods for exploring the chemical space by comparing the collaboration between human experimenters with an algorithm-based search against their performance individually to probe the self-assembly and crystallization of the polyoxometalate cluster $Na_6[Mo_{120}Ce_6O_{366}H_{12}(H_2O)_{78}]\cdot 200H_2O$ (**1**). We show that the robot-human teams are able to increase the prediction accuracy to 75.6 ± 1.8%, from 71.8 ± 0.3% with the algorithm alone and 66.3 ± 1.8% from only the human experimenters demonstrating that human-robot teams can beat robots or humans working alone.



INTRODUCTION

The scientific exploration of the vast chemical space for the discovery of new molecules has always been a challenging endeavor, since it is estimated that there are approximately 10^{60}−10^{100} synthetically feasible molecules.[1,2] As a result, the discovery of new chemical reactions can be a time-consuming process[3] especially when relying in traditional synthesis methods.[4] A huge improvement in reactivity prediction came with the development of computational methods (such as density functional theory, DFT, and empirical force field methods), which can screen a large number of candidate compounds in silico, reducing the need for all experiments to actually be carried out.[2,5] However, these methods can be computationally demanding as the system grows in complexity and are limited in that only ground-state structures can be calculated, ignoring metastable and transient species.[5] The emergence of artificial intelligence (AI) methods, and their implementation in chemistry, offers another avenue of exploration for chemical reactivity, see Figure 1. These methods have been facilitated by the availability of both big data[6,7] and open-source code for the training of algorithms.[8,9] A subfield of AI that has recently found applications in chemistry is machine learning,[5,10] which relies on data in order to construct a model of the chemical space under investigation. An advantage of machine learning is that the mechanistic details of the system do not need to be explicitly known in order to predict the probability of a given outcome or property of interest. Recent progress in automated chemistry[11] and online analysis[12,13] has allowed experimenters to build robots capable of exploring for chemical reactivity in a more autonomous way.[14,15] This means that robotic platforms can easily gather the data needed to implement machine learning algorithms. Nevertheless, the vast majority of algorithms (irrespective of their type) are not fully autonomous and require some guidance from the user, ranging from the choice of the internal parameters[16] (hyperparameters) for training to the selection of the algorithm for a specific chemical query[10] and the selection of the variables by the experimental scientist. Although deep-learning approaches (such as neural networks) have shown promise for predicting rules-of-thumb[17,18] in chemical synthesis, they suffer from two major drawbacks: First, they require large amounts of high-quality data in order to learn effectively, and second, they have difficulty operating outside their knowledge base.[10] In the former case, the problem stems from the fact that in chemistry, we are often limited to a relatively small number of high-quality data points (usually in the order of hundreds or thousands), while deep-learning methods are more attuned to systems with millions of data points (e.g., image recognition or text processing). In the latter case, machine learning methods can be predictive but not necessarily interpretable, since they tend to ignore molecular context because of the way that the data is represented in their model.[10] Therefore, the interaction with experimental scientists is important in order to assess these predictions,

Received: April 10, 2019
Published: April 26, 2019

Downloaded via UNIV OF GLASGOW on July 16, 2019 at 14:36:11 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

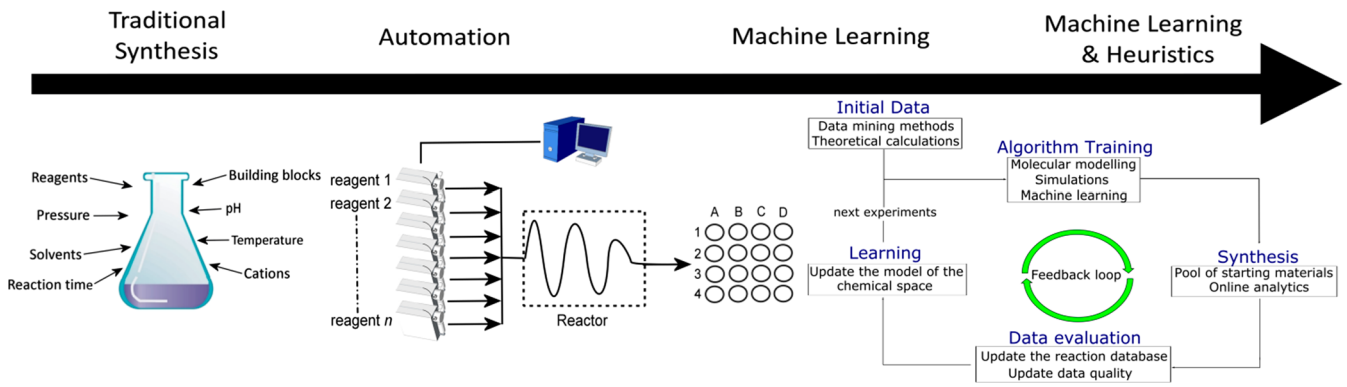


Figure 1. The evolution process from the traditional synthesis of one-pot methods to the automated high-throughput methods and more recently the advent of the increased use of machine learning methods for the exploration of the vast chemical space. For the case of the machine learning methods, in the beginning a hypothesis is formed starting from the available data in the literature and theoretical calculations. Then, this information is used to train an algorithm in recognizing patterns in the data and subsequently suggest a series of experiments to be performed. After the data evaluation, it is possible to update the model of the chemical space, and the cycle can begin again with training of the algorithm in the new acquired information.

and in the end, it is chemical intuition that determines which outcomes are valuable and which may be ignored.[5,10,19]

Intuition is generally described as heuristics, comprising strategies that human experimenters employ in problem solving and decision making by finding patterns, analogies, similarities, and rules-of-thumb in their data.[20,21] While automation has allowed generating, collecting, and storing data from scientific measurements in a very reliable and precise way, the field lacks uniform ways to process this data into concrete knowledge in the form of an analytical expression. The most significant advantage of intuition is that it does not require full information or knowledge of an unknown situation,[22−24] and in this way, it allows experimenters to perform well even in areas of high uncertainty.[25] Furthermore, the human mind is not able to process situations with a multitude of variables,[26] and as a result, it resorts to intuition and establishes a direction along which exploration can be performed in a consistent and meaningful way without getting lost in the details. In the context of chemistry, we can therefore only have a general overview of the system we are studying. Thus far, data mining methods are the closest approximations that have been developed as a means to substitute for human intuition in the experimental design.[10]

Within this framework, we propose that strategies based on chemists' intuition coupled with machine learning methodologies are a powerful alternative way to explore complex problems that involve large combinatorial spaces or nonlinear processes, where machine learning methods alone are unsuitable. Additionally, we propose that human intuition can help in guiding chemical synthesis, especially in cases where there is a lack of high-quality data. To our knowledge, there has been very little experimental work combining heuristics and machine learning methods. An algorithmic approach has been shown to detect nonlinear energy conservation laws without any prior knowledge of physics, kinematics, and geometry.[19] To achieve this, the algorithm automatically searched experimental motion-tracking data captured from various physical systems (ranging from simple harmonic oscillators to chaotic double-pendula) and built its own model of the physical space. Depending on the types of variables provided to the algorithm, different types of laws were derived. This dependence suggests that any analytical expression derived from a given computational method is amenable to human interpretation, and so close collaboration between the human factor and an algorithm can help in finding interesting phenomena more rapidly than before.

In the field of chemistry, Raccuglia et al.,[27] implemented machine learning algorithms to predict reaction outcomes of vanadium compounds by using data from unsuccessful and unreported syntheses (labeled by the authors as "dark" data) and compared the efficiency of the algorithms with the typical strategies that human chemists apply. Additionally, they demonstrated how the prediction accuracy of the model provided by the algorithm is higher than that of the human chemical intuition, both for single-crystalline and polycrystalline products. Nevertheless, the comparison is indirect, since the authors use unreported data from their lab books as a database for their analysis and did not actively compare the methodologies that human experimenters employ when searching the chemical space of a given compound.

In our previous work, we have demonstrated that we can push the envelope of both the synthesis and the crystallization process of a new polyoxometalate (POM) compound.[28] Our method is drawn from recent advances for active data acquisition in the field of machine learning, known as active learning. Active learning consists of methodologies that can decide what experiments should be performed next in order to improve the understanding of a system in the most efficient way. We studied how human experimenters approach the exploration and modeling of crystallization conditions for a given POM compound and directly compared the performance of their strategies to a machine learning approach. We hypothesized that this could be a first step to developing a new approach, which could combine the intuition of the chemists with machine learning in order to explore complex chemical systems and identify new phenomena. Additionally, in the work of Granda et al.,[15] and inspired by strategies based on chemists' intuition, it is demonstrated that a reaction system controlled by a machine learning algorithm is capable of exploring the space of chemical reactions quickly, especially if trained by an expert. An organic synthesis robot can perform chemical reactions and analyses faster than they can be performed manually as well as predict the reactivity of possible reagent combinations after conducting a small number of experiments, thus effectively navigating a chemical reaction space. By using real-time data from this robot, the predictions

DOI: 10.1021/acs.jcim.9b00304
J. Chem. Inf. Model. 2019, 59, 2664−2671

of reactivity are followed up manually by a chemist, leading to the discovery of four reactions.

Herein, we build on that previously reported work[28] of comparing an algorithm against the human intuition of human experimenters, and we attempt to combine them to explore the chemical space of the compound with chemical formula $Na_6[Mo_{120}Ce_6O_{366}H_{12}(H_2O)_{78}]\cdot 200H_2O$ (**1**) (hereafter also mentioned as $\{Mo_{120}Ce_6\}$). In this context, the key question is whether we can quantify the way soft knowledge (i.e., heuristics and more concretely human intuition) and hard knowledge (i.e., the increased computational capability of a machine learning method) interact with each other as a team and, potentially, gain some insights in how this collaboration works. Ultimately, we want to benefit from these insights and improve the way we explore the vast chemical space. In Figure 2, we illustrate in a simplified conceptual scheme of our observations from the evolution of the prediction accuracy as a result of our experiments previously done.[28] There are two areas of special interest for the performance of the combination of human intuition and machine learning that can be observed: area A and area B. In the case of the former, the resulting performance from the experiments is better than by simply utilizing an algorithm. In the case of the latter, the performance lies between that of the human experimenters and the algorithm.

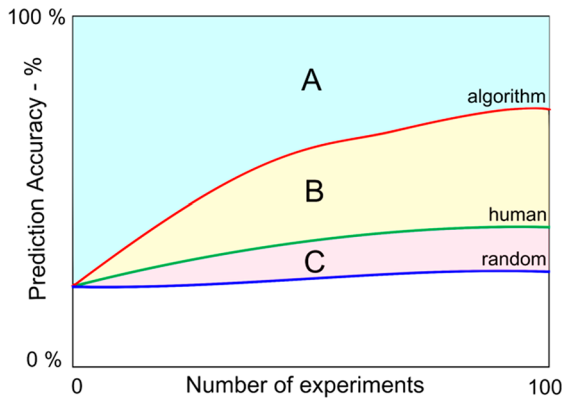


Figure 2. A conceptual scheme which represents the general trends of the evolution of the prediction accuracy based on our previous study.[28] In area A, the performances are higher than the ones observed from the algorithm by itself. In area B, the performances lie between that of the human experimenters and the algorithm. Lastly, in area C, the performances are only marginally better than a random search. Color scheme: algorithm, red line; human experimenters, green line; and random search, blue line.

Therefore, we aim to see how and if the team effort of human intuition and machine learning is able to increase its efficiency and lie in area A of Figure 2. The part of the machine learning is expressed with the use of an algorithm as described in the Supporting Information (SI), part 5. As for the reasons that we are interested in crystallization, the first is because of its broad implementation in the pharmaceutical industry and materials chemistry (e.g., with the isolation of new molecules that can be used as active pharmaceutical ingredients in drugs), and the second is because the crystal structure presents some inherent challenges as a result of the difficulty to find a format able to represent a crystalline solid in such a way so that it can be easily fed to a statistical learning procedure. We believe that finding a way to digitalize intuition can have an impact on both accelerating the discovery rate of new phenomena in more complex systems and on how young chemists are trained, since we can distill a vast body of seemingly random chemical information into an organized and interconnected web of knowledge.[20,28,29]

METHODS

In our previous work,[28] we observed how the models computed by an algorithm are able to improve their prediction accuracy better than the models suggested by the human experimenters (82.4% over 77.1%, respectively, with a baseline performance of 68.1%). The algorithm we implemented is a classifier assigning labels (e.g., crystal/no-crystal) to regions of a parameters space; the human experimenters were volunteer Ph.D. students in our group familiar with inorganic chemistry synthesis; and the baseline method used as a control was a random search, rendering it blind to both the initial and the subsequently collected data. As a result, this difference in

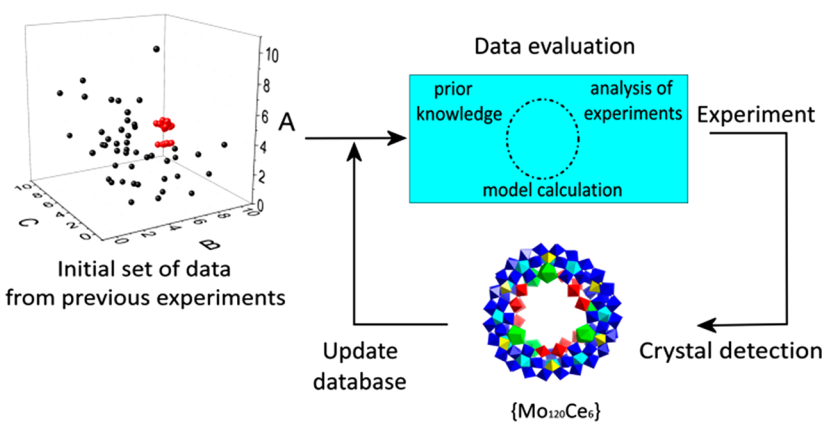


Figure 3. Experimental protocol describing the decision-making process during the exploration of the chemical space of $\{Mo_{120}Ce_6\}$ that was implemented during this work. An initial set of data serves as a starting point for experiments to perform next after analysis and model calculation. The experiments are performed in a fully automated platform, and the outcome is observed and recorded in an updated version of the initial database. Coloring code of the building units found in $\{Mo_{120}Ce_6\}$: $\{Mo_2\}$, red; $\{Mo_1\}$, yellow; $\{Mo_8\}$, blue with central atom in cyan; Ce, green. For the 3D plot of the initial set of data, the axes represent the following: A, $Na_2MoO_4\cdot 2H_2O$ 1 M and $Ce(NO_3)_3\cdot 6H_2O$ 0.1 M (mL); B, $HClO_4$ 1 M (mL); and C, $NH_2NH_2\cdot 2HCl$ 0.25 M (mL).

DOI: 10.1021/acs.jcim.9b00304
J. Chem. Inf. Model. 2019, 59, 2664−2671

performance between algorithm and human experimenters indicates the effect that the different strategies followed by the human experimenters can have when they are based solely on their intuition.

The basis for the current work is the formation and crystallization of cluster (**1**), and the general experimental procedure is depicted in Figure 3, where teams are formed consisting of human experimenters and an algorithm with the objective to explore chemical space together. In order to start their exploration, these teams are provided with the experimental conditions for the formation of (**1**): first, the chemicals involved in its synthesis (SI, part 2); second, the experimental protocol for the synthesis and crystallization process; and third, an initial set of data consisting of successful and unsuccessful crystallization experiments (SI, part 3).

The flowchart of Figure 4 describes the decision-making process of the human experimenter and algorithm teams. In

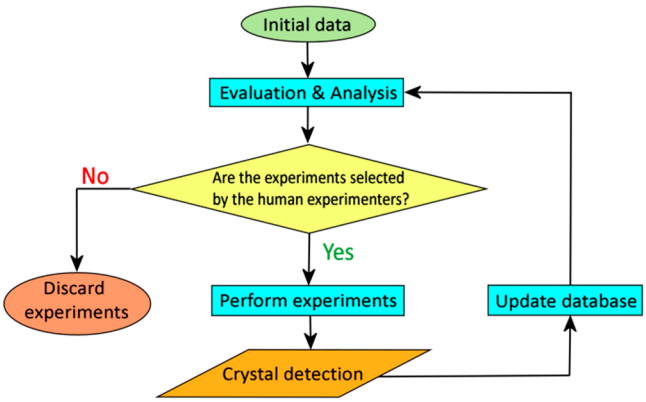


Figure 4. Decision-making process of the exploration protocol for the collaboration between the algorithm and the human experimenter. Starting from an initial set of data, the algorithm evaluates these results and suggests 20 experiments. For the next stage, the human experimenters select 10 experiments to run in the platform. The other 10 experiments that are not selected are discarded. After the reaction is finished, we wait for the crystallization time, and the database of experiments is updated with the outcome of the reaction before starting again, giving a loop that is repeated 10 times.

the beginning, the initial set of data (SI, part 3 and Table S2) serves as the starting information used to decide what experiments to perform next. As a first step, the algorithm evaluates these experiments and builds a model of the chemical space. Based on that model, the algorithm provides us with a list of 20 suggested experiments. Then, these experiments are presented to the human experimenters, and they select 10 out of them to perform in the platform. Finally, we perform the experiments in an automated platform (SI, part 4) and receive the information about the presence or the absence of crystals for each of the requested experiments by illuminating the samples under a strong white light-emitting diode (3300−3500 lux at a distance of 5 cm). The process is repeated 10 times per method for a total of 100 experiments each. At each iteration, all data collected previously are integrated in the decision process for generating the next set of 20 experiments.

At this point, we should mention that our previous experience with this chemical system allowed us to perform 10 experiments per day and wait overnight for crystallization of the product.[28] For the investigations described herein, we needed to modify our experimental procedure, as shown in Figure 4, in order to accommodate for the addition of the intuition of the human experimenters (through their suggestions) as an additional factor in the decision making of the algorithm. Therefore, the algorithm was altered in order to produce a list of 20 suggested experiments, and the human experimenters were instructed to select 10 to be performed in the platform. The machine learning parameters of the algorithm used a 10-fold cross-validation to search the best C and γ hyperparameters, where C is the regularization parameter and γ is the kernel coefficient of the radial basis function (ESI, part 6). To do this, we ran a cross-validation with all possible combinations of C and γ within the set (10^{-5}, $10^{-4.5}$, 10^{-4}, ..., $10^{4.5}$, 10^{5}) and selected the C and γ values producing the smallest classification error, that is, the most accurate model. In our case, these values are $C = 100$ and $\gamma = 10^{-3/2}$, and they are the same ones as before[28] since both are extremely important in order to tune the model provided by the algorithm. In the case where different values were used, it is possible to get entirely different performance from the algorithm,[16] and this means we are unable to directly compare the results across methods. To do this work, we use scikit-learn,[30,31] a machine learning library built for Python.

RESULTS AND DISCUSSION

The data from the experiments, unless presented as an average result over multiple runs, are represented as H1 and H2 for the human experimenters, A1 and A2 for the algorithm runs, R1 and R2 for the random search, and T1, T2, and T3 for the teams of human experimenters and algorithm. The results shown are after the end of the 100 experiments requested at the beginning of our study. The exploration performed by all methods is quantified by using metrics such as the evolution of the prediction accuracy, the similarity of experiments, and the volume exploration. A brief theoretical background behind these metrics is provided in the SI, parts 8.1−8.4.

Prediction Accuracy. The evolution of the prediction accuracy of each method trained on the data collected in each run can be seen in Figure 5. The quality of the prediction (i.e., the percentage of time a crystal prediction is accurate) is expected to increase as more data are collected. The initial

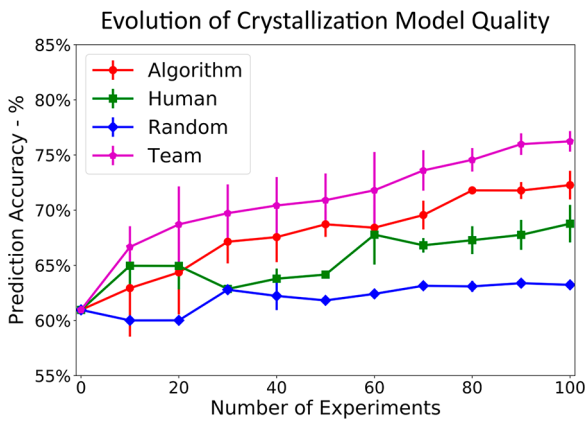


Figure 5. Average of the prediction accuracies for all methods with error bars as implemented by RandomForest. We can observe the higher variability in the error of the team in comparison to the other methods, which can be attributed to the different methodologies that the human experimenters followed for their calculations during the exploration of the chemical space (see ESI, part 8.3, Table S4 and Figure S24).

DOI: 10.1021/acs.jcim.9b00304
J. Chem. Inf. Model. 2019, 59, 2664−2671

prediction quality based on the initial set of data provided to all methods is 66.5%. We can observe that the team was able to collect better quality data than the algorithm (75.6 ± 1.8% over 71.8 ± 0.3%) and improved its classification accuracy the most. Since we used 989 experimental points (ESI, part 8.5, Tables S5 and S6) in order to compute the prediction quality, this means that a 3.8% difference represents on average 38 additional experiments correctly predicted in our data set. This difference is quite substantial both in terms of our model and in machine learning grounds.

In light of Figure 5, we can observe that the interaction of the human intuition and the algorithm manages to increase the performance of the individual parts and achieve higher efficiencies than the algorithm by itself. As for the existence of the larger variability of the standard deviation for the teams, we can only assume at this stage that it is the result of the different methodologies from the human experimenters in their interaction with the algorithm (see ESI, part 9).

Similarity of Experiments. For this metric, we calculate how many other experiments lie within a specific radius R (we use $R = 2$) in the parameters' space (see ESI, part 8.4, Figure S26). This distance is a similarity measure between experiments: A large value indicates similar experiments, while a small value indicates more explored chemical space. In Figure 6, we plot this similarity metric as more experiments are performed. First, we note that in the initial set, 95% of the experiments leading to crystals are within a radius of $R = 2$ of each other in the chemical space.

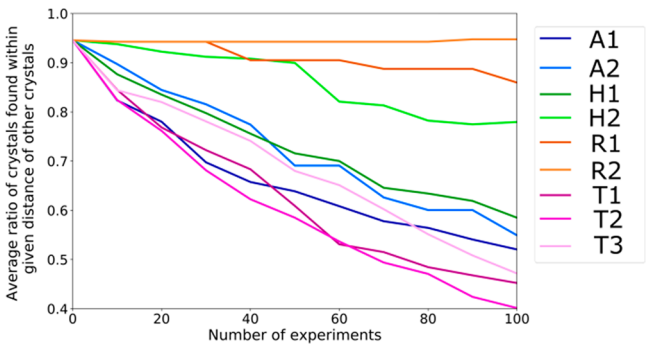


Figure 6. Similarity metric of the experiments plotted as a comparison of the average ratio of crystals found within a given distance of other crystals. The faster this ratio drops, the wider the exploration. The data are represented as H1 and H2 for the human experimenters, A1 and A2 for the algorithm runs, R1 and R2 for the random search, and T1, T2, and T3 for the teams of human experimenters and algorithm. Note how two different groups emerge from this data: The first group consists of H2, R1, and R2, and the second group consists of H1, A1, A2, T1, T2, and T3.

We can observe that T1, T2, and T3 reduce this ratio faster than any other method, indicating a wider exploration and thus less data points in the vicinity of each other. A similar dynamic can be observed in the algorithm runs A1 and A2 which follow the same trend of fast exploration as the teams. As for the random search (R1 and R2), there is no improvement in their performance. In the case of the human experimenters H1 and H2, we can observe two different behaviors: H1, who shares the same trend as the algorithm and the teams, and H2, who is closer to the random search (R1 and R2). We have previously demonstrated[29] that this broad difference between these runs can be attributed to conservative strategies of exploration, where small steps of exploration are performed, that can limit the information that we can obtain about the chemical landscape. At this stage, we can make a ranking of which individual run is better in exploring the chemical space: T2 > T1 > T3 > A1 > A2 > H1 ≫ H2 > R2 > R1. This ranking resonates with the observations we made with the previous metric in Figure 5 and is the first clear evidence that the collaboration of the algorithm and the human intuition can perform better than each of these two parts individually. In light of this metric, we hypothesize that the effect of the different strategies adopted by the human experimenters as well as their inherent biases can be mitigated with the collaboration of the machine learning methods in exploring the chemical space.

Volume Exploration. Considering the crystallization area as a proxy for the volume of the parameters' space of the chemicals involved in the experiments, a valuable metric is to estimate how much of the crystallization volume has been explored by each method. Following the results from our experiments, we plotted the average explored volume as a function of the number of experiments performed, see Figure 7. For the volume calculation, the volume of the experiments leading to crystals was computed (ESI, part 8.2).

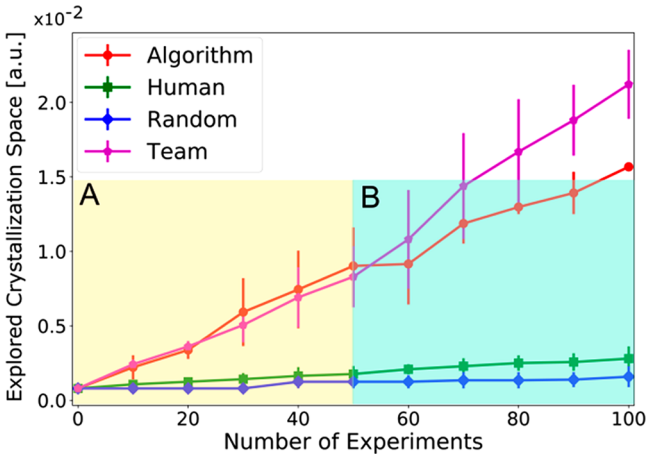


Figure 7. Average explored volume of the crystallization space by the four methods (algorithm, human experimenter, random, and team) along with their respective error bars. We can observe two areas of interest between algorithm and team: area A (from the beginning until experiment 50), where they follow a similar behavior, and area B (experiments 50−100), where they start deviating from each other and become distinct following a different path. The respective values of volume and standard deviation are presented in the SI, part 8.2, Table S3 and Figure S23.

The y-axis in the results of Figure 7 corresponds to a four-dimensional volume of all crystal points in the parameter space of the chemical reagents. Since each parameter is in mL units, the y-axis unit is strictly speaking mL^4, which has no intuitive meaning, and therefore the results should be interpreted relative to each other as arbitrary units (a.u.) and not as absolute values. The error bars of the standard deviation depict the effect that the different methodologies can have in collecting useful data for improving the calculated model of the chemical space in each iteration. The respective values of volume and standard deviation for each individual run are provided in Table S3. From the values in this table, we can observe that from the sixth run onward, the team increases the

DOI: 10.1021/acs.jcim.9b00304
J. Chem. Inf. Model. 2019, 59, 2664−2671

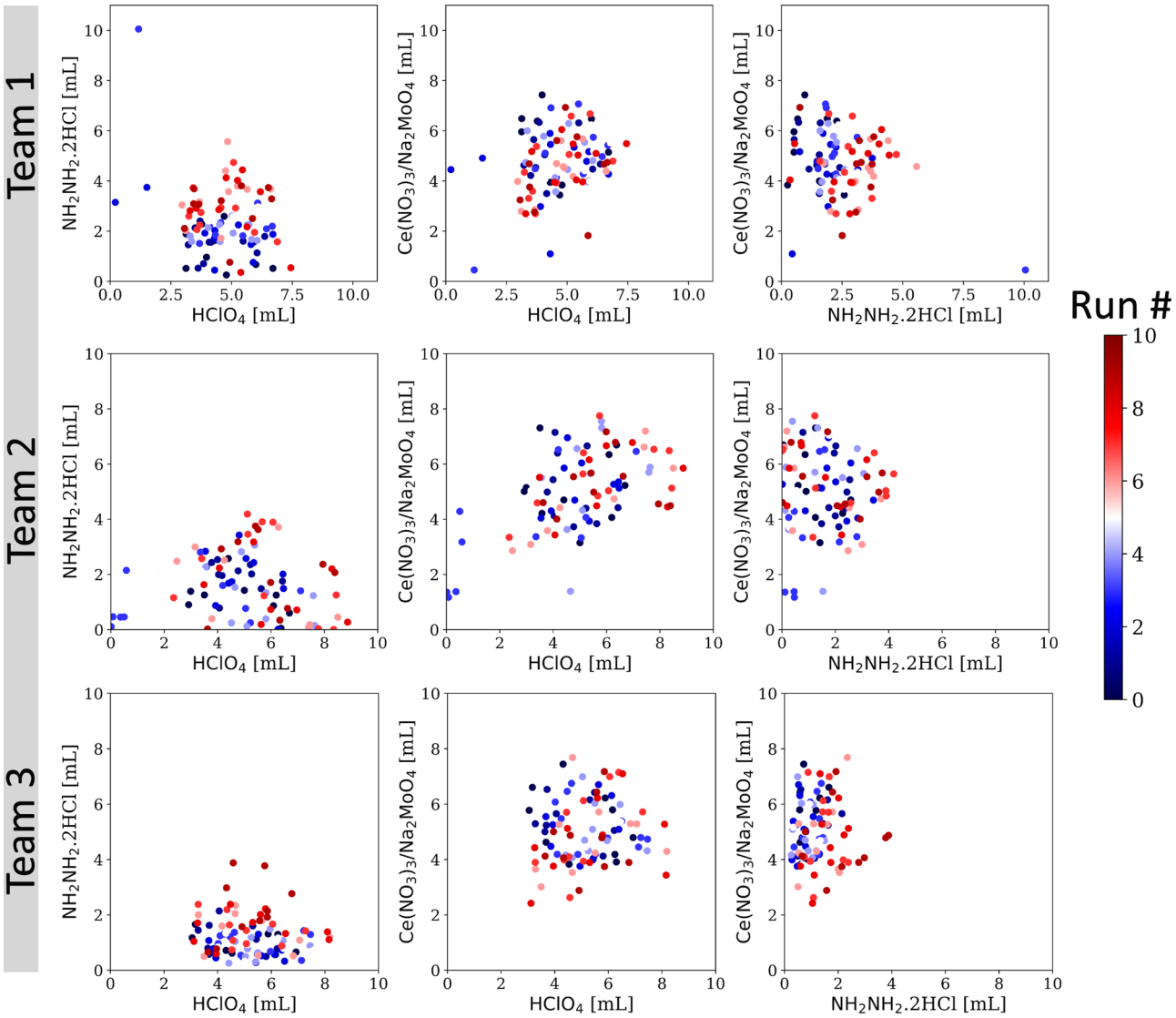


Figure 8. 2D plot represents the choice of the selected experiments during the different runs of the teams. The data from the different runs are plotted against the same surface, and the distinction is made by using a color scheme. The first run is represented in dark blue, while the last in dark red. For Team 1 (T1), we can observe the strong preference for experiments with an increased amount of reducing agent ($NH_2NH_2{\cdot}2HCl$) over the course of the experiments (graphs in the left and right). For Team 2 (T2), the middle graph shows how the increasing amounts of perchloric acid are selected during this study as also reported in the experimental protocol. For team 3 (T3), notice the similarity of behavior in relation to T1.

amount of space substantially it covers (from 1.08×10^{-2} a.u. to 2.12×10^{-2} a.u.), while the algorithm exhibits a relatively slower pace of exploration (from 0.91×10^{-2} a.u. to 1.56×10^{-2} a.u.).

The difference between algorithm and human experimenters can be explained by the fact that the algorithm is agnostic to the chemical environment and untied to prior chemical knowledge. This way it can perform jumps in the chemical space straight into the believed boundaries between crystal and no-crystal. On the contrary, human experimenters have drastically varied strategies depending on personal perceptions and biases of the particular chemistry involved in the system under study. A noticeable feature of Figure 7 is that the collaboration of the human experimenter and an algorithm seem to lift this difference between the two and the team work allows for more chemical space to be covered in the same amount of time. Furthermore, the team work manages to outperform the algorithm despite the differences of the exploring strategies followed by the human experimenters.

Interaction between Human Experimenters and Algorithm. We also attempted to understand the interaction between the human experimenter and the algorithm in a deeper level by depicting this interaction as a two-dimensional (2D) contour plot of the experiments over the different generations, see Figure 8. We observe that T1 is primarily focused on the amount of reducing agent ($NH_2NH_2{\cdot}2HCl$) and perchloric acid ($HClO_4$). In terms of the use of the reducing agent, we can observe a direction toward areas of higher amounts (left graph). Although the experimenter reports the use of perchloric acid as important for their protocol (SI, part 9, Team 1), it is not as evident from this plot since the selection of perchloric acid appears to be evenly distributed (middle graph). Additionally, the amounts of perchloric acid that are used for the experiments remain constrained between 2.5 and 7.5 mL. Another feature that we notice is the decrease of the ratio of $Na_2MoO_4{\cdot}2H_2O$/$Ce(NO_3)_3{\cdot}6H_2O$ during the study (right graph). Given the provided experimental protocol, we are not able to comment

DOI: 10.1021/acs.jcim.9b00304
J. Chem. Inf. Model. 2019, 59, 2664−2671

whether this is a feature that was also taken into account from the beginning but was not described, or if it occurred because of the specific selection of the experimental variables (i.e., the reducing agent and the perchloric acid) as guides for the exploration.

In the case of T2, the reported key guides for the exploration are the amount of reducing agent and the ratio of Mo/Ce (see SI, part 9, Team 2). In Figure 8, we can observe a more widespread search in terms of the reducing agent (graphs in left and right). Furthermore, we can observe a tendency to use more perchloric acid (middle graph), as it has also been reported in their experimental protocol. As for the ratio of Mo/Ce, it seems to be decreasing but still remaining in a region between 4 and 8 mL (middle and right graph). Finally, for the case of T3, there seems to be a lot of similarities with T1 in the direction of the reagents, although the reported experimental guide is only the ratio of Mo/Ce (SI, part 9, Team 3). A reason behind the choice of these common variables by the human experimenters is that small amounts of perchloric acid will not provide a low enough pH in order to reduce the system, whereas excessive amounts of $NH_2NH_2{\cdot}2HCl$ will cause overreduction. On the other hand, a small ratio of Mo/Ce will cause a deficiency in Mo, and the wheel will not be able to form. The trends that we can observe in the nonselected experiments of Figure 4 mirror the reasoning of the human experimenters, as described in their protocols, and allow us as a whole to derive preferred directions in the experimental procedure as well as identify the specific variables used for the exploration of the chemical space of $\{Mo_{120}Ce_6\}$ by identifying patterns in the experimental data. Nevertheless, it is not possible to directly unveil the trends that we observed in Figures 5−7 since Figure 8 is only a qualitative perspective of the data.

CONCLUSIONS

In our previous study[28] we hypothesized that the combination of both machine learning and intuition could be the first step to developing a new approach in order to explore complex chemical systems. This work demonstrates the significant impact that collaboration between human and machine can have, as a significantly higher performance is achieved by working together than either the algorithm or human experimenter could achieve individually. The most important advantage of intuition is its ability to perform well even in areas of high uncertainty. One such area is the lack of high-quality data in chemistry, and this is the framework around which this work was developed. The increased computational power of a machine learning model can allow us to identify hidden patterns in the data, while the human intuition can develop the direction for the experiments. In this way, the inherent personal and chemical biases of the human experimenter can be mitigated, and more "adventurous" studies of large combinatorial spaces or nonlinear processes can be accomplished. We were able to observe and quantify the effects of the team work between human and machine, but not without many problems arising from the different ways in which experimental procedures are documented. This reinforces the imperative need to find a way to digitize this knowledge. We believe that machine learning methods should be viewed as tools in order to assist human experimenters rather than replace them, and these results provide a proof of concept of how this interaction can work. There is a lot more ground to cover in this area, but we feel that bringing together advanced machine learning with human intuition will be transformative and lead to new methodologies in exploring complex problems.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.9b00304.

Full analytical data and a description of the analytical methods is given in the PDF (PDF)

The data collected for the analysis is presented in the ZIP file experimental data (ZIP)

The code used to analyse the data given in the experimental zip file is in the ZIP file code (ZIP)

AUTHOR INFORMATION

Corresponding Author

*E-mail: Lee.Cronin@glasgow.ac.uk.

ORCID

Vasilios Duros: 0000-0002-8373-4362
Jonathan Grizou: 0000-0002-2211-4389
S. Hessam M. Mehr: 0000-0001-7710-3102
Andrius Bubliauskas: 0000-0001-8680-1914
Przemysław Frei: 0000-0003-4688-8247
Haralampos N. Miras: 0000-0002-0086-5173
Leroy Cronin: 0000-0001-8035-5757

Author Contributions

The idea was conceived by L.C. and developed by V.D. and J.G. V.D. did all the automated work with data analysis with help from J.G., A.S., and H.M. The algorithm was developed by J.G., V.D., and L.C. Synthesis was done using the robot by V.D. and the humans who talked were A.B. and P.F. H.N.M. helped with data analysis. All the authors helped write the manuscript.

Funding

We gratefully acknowledge financial support from the EPSRC (grant nos EP/P00153X/1, EP/J015156/1, EP/K021966/1, EP/K038885/1, EP/L015668/1, and EP/L023652/1) and ERC (project 670467 SMART-POM).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We would like to thank Dr. Juan Manuel Parrilla Gutiérrez for reading the manuscript and making important comments.

ABBREVIATIONS

AI, artificial intelligence; POMs, polyoxometalates; a.u., arbitrary unit; 2D, two-dimensional

REFERENCES

(1) Lipinski, C.; Hopkins, A. Navigating Chemical Space for Biology and Medicine. *Nature* **2004**, *432*, 855−861.

(2) Schneider, G.; Fechner, U. Computer-Based *de novo* Design of Drug-Like Molecules. *Nat. Rev. Drug Discovery* **2005**, *4*, 649−663.

(3) Collins, K. D.; Gensch, T.; Glorius, F. Contemporary Screening Approaches to Reaction Discovery and Development. *Nat. Chem.* **2014**, *6*, 859−871.

(4) Trobe, M.; Burke, M. D. The Molecular Industrial Revolution: Automated Synthesis of Small Molecules. *Angew. Chem., Int. Ed.* **2018**, *57*, 4192−4214.

DOI: 10.1021/acs.jcim.9b00304
J. Chem. Inf. Model. 2019, 59, 2664−2671

(5) Graser, J.; Kauwe, S. K.; Sparks, T. D. Machine Learning and Energy Minimization Approaches for Crystal Structure Predictions: A Review and New Horizons. *Chem. Mater.* **2018**, *30*, 3601−3612.
(6) Tetko, I. V.; Engkvist, O.; Koch, U.; Reymond, J. L.; Chen, H. BIGCHEM: Challenges and Opportunities for Big Data Analysis in Chemistry. *Mol. Inf.* **2016**, *35*, 615−621.
(7) Kalinin, S. V.; Sumpter, B. G.; Archibald, R. K. Big-Deep-Smart Data in Imaging for Guiding Materials Design. *Nat. Mater.* **2015**, *14*, 973−980.
(8) Steinbeck, C.; Han, Y.; Kuhn, S.; Horlacher, O.; Luttmann, E.; Willighagen, E. The Chemistry Development Kit (CDK): An Open-Source Java Library for Chemo- and Bioinformatics. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 493−500.
(9) O'Boyle, N. M.; Guha, R.; Willighagen, E.; Adams, S. E.; Alvarsson, J.; Bradley, J. C.; Filippov, I. V.; Hanson, R. M.; Hanwell, M. D.; Hutchison, G. R.; James, C. A.; Jeliazkova, N.; Lang, A. S.; Langner, K. M.; Lonie, D. C.; Lowe, D. M.; Pansanel, J.; Pavlov, D.; Spjuth, O.; Steinbeck, C.; Tenderholt, A. L.; Theisen, A. L.; Murray-Rust, P. Open Data, Open Source and Open Standards in Chemistry: The Blue Obelisk Five Years On. *J. Cheminf.* **2011**, *3*, 37.
(10) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine Learning for Molecular and Materials Science. *Nature* **2018**, *559*, 547−555.
(11) Schneider, G. Automating Drug Discovery. *Nat. Rev. Drug Discovery* **2017**, *17*, 97−113.
(12) Houben, C.; Lapkin, A. A. Automatic Discovery and Optimization of Chemical Processes. *Curr. Opin. Chem. Eng.* **2015**, *9*, 1−7.
(13) Sans, V.; Cronin, L. Towards Dial-a-Molecule by Integrating Continuous Flow, Analytics and Self-Optimisation. *Chem. Soc. Rev.* **2016**, *45*, 2032−2043.
(14) Dragone, V.; Sans, V.; Henson, A. B.; Granda, J. M.; Cronin, L. An Autonomous Organic Reaction Search Engine for Chemical Reactivity. *Nat. Commun.* **2017**, *8*, 15733.
(15) Granda, J. M.; Donina, L.; Dragone, V.; Long, D.-L.; Cronin, L. Controlling an Organic Synthesis Robot with Machine Learning to Search for New Reactivity. *Nature* **2018**, *559*, 377−381.
(16) Wang, B.; Gong, N. Z. Stealing Hyperparameters in Machine Learning. Proceedings from the *2018 IEEE Symposium on Security and Privacy (SP)*, San Francisco, CA, May 20-24, 2018; IEEE: Piscataway, NJ, 2018; pp 629−645.
(17) Segler, M. H. S.; Waller, M. P. Neural-Symbolic Machine Learning for Retrosynthesis and Reaction Prediction. *Chem. - Eur. J.* **2017**, *23*, 5966−5971.
(18) Segler, M. H. S.; Preuss, M.; Waller, M. P. Planning Chemical Syntheses with Deep Neural Networks and Symbolic AI. *Nature* **2018**, *555*, 604−610.
(19) Schmidt, M.; Lipson, H. Distilling Free-Form Natural Laws from Experimental Data. *Science* **2009**, *324*, 81−85.
(20) Graulich, N.; Hopf, H.; Schreiner, P. R. Heuristic Thinking Makes a Chemist Smart. *Chem. Soc. Rev.* **2010**, *39*, 1503−1512.
(21) Gibb, B. C. Chemical Intuition Or Chemical Institution? *Nat. Chem.* **2012**, *4*, 237−238.
(22) Evans, J. S. B. T. Dual-Processing Accounts of Reasoning, Judgment. *Annu. Rev. Psychol.* **2008**, *59*, 255−278.
(23) Gigerenzer, G.; Gaissmaier, W. Heuristic Decision Making. *Annu. Rev. Psychol.* **2011**, *62*, 451−482.
(24) Kahneman, D.; Klein, G. Conditions for Intuitive Expertise: A Failure to Disagree. *Am. Psychol.* **2009**, *64*, 515−526.
(25) Tversky, A.; Kahneman, D. Judgment under Uncertainty: Heuristics and Biases. *Science* **1974**, *185*, 1124−1131.
(26) Halford, G. S.; Baker, R.; McCredden, J. E.; Bain, J. D. How Many Variables Can Humans Process? *Psychol. Sci.* **2005**, *16*, 70−76.
(27) Raccuglia, P.; Elbert, K. C.; Adler, P. D.; Falk, C.; Wenny, M. B.; Mollo, A.; Zeller, M.; Friedler, S. A.; Schrier, J.; Norquist, A. J. Machine-Learning-Assisted Materials Discovery Using Failed Experiments. *Nature* **2016**, *533*, 73−76.
(28) Duros, V.; Grizou, J.; Xuan, W.; Hosni, Z.; Long, D.-L.; Miras, H. N.; Cronin, L. Human versus Robots in the Discovery and Crystallization of Gigantic Polyoxometalates. *Angew. Chem., Int. Ed.* **2017**, *56*, 10815−10820.
(29) Kovács, F.; Legány, C.; Babos, A. Cluster Validity Measurement Techniques. Proceedings from the *5th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases*, Madrid, Spain, February 15−17, 2006; World Scientific and Engineering Academy and Society (WSEAS): Stevens Point, WI, 2006; pp 1−11.
(30) Talanquer, V. Chemistry Education: Ten Heuristics To Tame. *J. Chem. Educ.* **2014**, *91*, 1091−1097.
(31) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; duchesnay, E. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2010**, *12*, 2825−2830.

DOI: 10.1021/acs.jcim.9b00304
J. Chem. Inf. Model. 2019, 59, 2664−2671