Title

Taking model pursuit seriously

Author Information

- **Name** | HyeJeong Han
- **Affiliation** | Graduate School of Science and Technology Policy, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea
- **E-mail** | hhj29@kaist.ac.kr
- **ORCID** | 0000-0003-1288-7705

Abstract

This paper aims to develop an account of the pursuitworthiness of *models* based on a view of models as epistemic tools. This paper is motivated by the historical question of why, in the 1960s, when many scientists hardly found QSAR models attractive, some pharmaceutical scientists pursued Quantitative Structure-Activity Relationship (QSAR) models despite the lack of potential for theoretical development or empirical success. This paper addresses this question by focusing on how models perform their heuristic functions as epistemic tools rather than as potential theories. I argue that models perform their heuristic function by "constructing" phenomena from data in the sense that they allow the model users who interact with the medium of the models to recognise the phenomena *as such*. The constructed phenomena assist model users in identifying which conditional hypotheses that are focused on low-level regularities concerning entities such as chemical compounds are more "testworthy," a concept that links the costs associated with hypothesis testing with the fertility of the hypothesis.

Keywords

Model pursuit, Epistemic tools, Phenomena construction, Testworthiness, QSAR, Drug design

Acknowledgements

I especially thank Grant Fisher for his invaluable guidance and helpful suggestions for writing and improving this paper. I also thank Dunja Šešelja and two anonymous reviewers for their constructive comments on the earlier

draft of this paper.

Declarations

Funding: Not applicable.

Competing interests: Not applicable.

Ethical approval: Not applicable.

Informed consent: Not applicable.

1 Introduction

Since Larry Laudan's (1977) context distinction between pursuit and acceptance, there has been an increase in philosophical interest in the topic of scientific pursuitworthiness.[1] This rise in interest has been largely fuelled by the desire to analyse aspects of science that have been neglected by the logical empiricist tradition and thus better capture the actual practices of science.[2] However, some features of the current pursuit discussion make us wonder whether the goals have been met. The first notable feature is that, while theories have been at the centre of the debate, *models* have been treated as mere theoretical representations with little regard for their distinct roles from theories in scientific practices. Although some philosophers have discussed model pursuit, they have reduced it to or blurred it with theory pursuit (e.g., French, 1995, 1997; Šešelja & Straßer, 2014; Nyrup, 2020; Haueis & Kästner, 2022). Furthermore, despite the widespread recognition in the pursuit literature of the link between practical and epistemic considerations (e.g., Šešelja et al., 2012; DiMarco & Khalifa, 2022), the practical dimensions have frequently been left out of the epistemic analysis, being regarded as those external to scientific reasoning.[3] One aspect exemplifying this feature is that case studies in the pursuit literature have largely been restricted to publicly-funded academic research, missing out on industry-funded research, which accounts for a sizable portion of today's science.[4]

A motivation for writing this paper is to bring the pursuit discussion closer to actual scientific practices by addressing the issues raised above. This paper, in particular, aims to develop an account of model pursuit based on the view of models as epistemic tools. The epistemic tool view refers to a recently developed view of models that considers models as material artefacts constructed to serve a variety of specific purposes (e.g., Knuuttila, 2011; Knuuttila & Boon, 2011; Currie, 2017). As one of the views that highlight the partial independence of models from theories, and as a view that incorporates material and practical dimensions into philosophical

---

[1] For example, an article collection titled "Pursuitworthiness in Scientific Inquiry" will be published by *Studies in History and Philosophy of Science* in 2023, edited by Jamie Shaw and Dunja Šešelja; the articles of the collection have been released since 2022.

[2] The most recent papers published in the 2010s and 2020s do not respond directly to the logical empiricists. Most of them, however, discuss scientific pursuit in light of Laudan and others' criticisms of the distinction between discovery and justification.

[3] Shan's (2020) account of "promisingness" is one of the few philosophical works that analyse the practical dimensions. Section 4.1 provides additional details about this account.

[4] In the OECD, for example, industry accounts for around 70% of R&D expenditures (OECD, 2021).

discussions on scientific models, this view can be used fruitfully in the analysis of model pursuit in actual scientific practices. Currently, in the pursuit literature, the dominant way of assessing which models are worth pursuing is based on their heuristic potential for theoretical development. However, by viewing models as epistemic tools, I will shift the focus from the models' theoretical development to the models' assistance for scientists in the assessment of very focused conditional hypotheses about entities such as chemical compounds, particularly in assessing which hypotheses are more worth testing.

To develop the argument, I will look into a historical case in a field that has been frequently overlooked in philosophical analysis despite its growing importance in actual scientific practices: chemical and biological research funded by industry. In particular, the pursuit of Quantitative Structure-Activity Relationship (QSAR) models in the pharmaceutical industry in the 1960s will be a good historical episode that shows how models perform this type of heuristic function. QSAR is a statistical modelling method that calculates the biological activity of a compound based on its chemical structure. Some pharmaceutical scientists in the 1960s pursued QSAR models in the hope that they would aid in the efficient development of new drugs. Looking back on this episode from today's perspective, it might be easy to conclude that QSAR models were indeed worth pursuing, because we already know that QSAR models have significantly advanced (see, e.g., Selassie & Verma, 2010) and delivered various benefits since then. In the 1960s, however, when QSAR models had not yet demonstrated empirical success in drug development nor offered any potential for theoretical development, many scientists did not find QSAR models appealing.[5] Herein lies the historical question of why some scientists invested time and resources in pursuing, i.e., working further with, QSAR models in the 1960s, despite the models' limitations at the time.

A key to solving this historical question is to examine how pharmaceutical scientists used QSAR models as epistemic tools to assess hypotheses about drug candidates, rather than as potential theories. In this paper, I will argue that models perform the heuristic function by "constructing" phenomena from data in the sense that they allow model users to recognise the phenomena *as such*. Importantly, when discussing the heuristic function of

[5] Given certain reports referencing Hansch and QSAR's low reputation in the 1960s—for example, a Merck researcher deemed Hansch's work "ridiculous" in 1965 (Hansch, 2003, p. 621), and no one at Abbott knew about Hansch's work until 1967 (Martin, 2018, p. 817)—it is plausible to say that it was not until the mid-1970s that Hansch and his QSAR became widely recognised as promising. A historical overview of QSAR models is provided in Section 3.

models, the accuracy of (the description of) the constructed phenomena is not important. Instead, what matters is that the constructed phenomena assist model users in identifying what I call the "testworthiness" of hypotheses. To clarify, the hypotheses in this paper do not refer to high-level theories, but rather to very focused conditional hypotheses about low-level regularities.[6] When a hypothesis is said to be testworthy, I mean that it is worthwhile to invest time and resources in the practices associated with the testing of the hypothesis. The assessment of the testworthiness of hypotheses is critical for the acquisition of new knowledge because if more testworthy hypotheses cannot be identified, future knowledge acquisition becomes very inefficient and even impossible given the limited resources in actual scientific practices. In this manner, models can assist model users in efficiently acquiring knowledge in the future.

This paper will proceed as follows. I will review the existing model pursuit literature in Section 2 and argue that the current approach is theory-centric. Section 3 will provide a historical overview of drug design and QSAR models, and Section 4 will revise the concept of the heuristic function of models to make it more suitable for model pursuit by viewing models as epistemic tools. In particular, I will claim that models can help model users obtain knowledge efficiently by assisting in determining the "testworthiness" of hypotheses, which refers to identifying whether it is worth allocating resources to testing very focused conditional hypotheses. Following that, in Section 5, I will state that models can "construct" phenomena from data in the sense that making the model users recognise the phenomena as such. Section 6 will articulate the relationship between the models' heuristic function and the construction of phenomena.

2 The heuristic function of models

While Laudan (1977) emphasised the demonstrated rate of problem-solving success as a criterion for pursuitworthiness, many other recent philosophers have emphasised forward-looking criteria, particularly those concerning heuristics. Among the recent literature on heuristics, Steven French's (1995, 1997) account of "heuristic fruitfulness" and Šešelja and Straßer's (2014) account of "programmatic character" are particularly relevant to model pursuit. In this section, I examine the existing literature on model pursuit, with a particular emphasis on these two accounts. Then I argue that they are theory-centric because, first, they emphasise the

[6] For more on the distinction between high-level theories and low-level regularities, see, e.g., Hacking (1983).

theoretical development of models and, second, they derive the source of heuristics for models from connections to other theories.

Tracing back the history of the philosophy of science, it has not been very long since models have been considered a serious topic of discussion. In the first half of the twentieth century, the so-called "received view" encouraged philosophers to focus solely on theories rather than models. Models were typically viewed as playing only subsidiary roles in relation to theories, such as providing an interpretation of a theory or heuristics for theoretical development.

One of the first attempts to take models seriously was to look at their analogical roles. Mary Hesse's (1970) tripartite distinction between analogies is a well-known example. She categorises analogies as positive, negative, and neutral based on whether the model and its target share certain properties. For example, in gas molecules and billiard ball models, the motion of billiard balls corresponds to the positive analogy, because gas molecules and billiard balls share these properties. By contrast, certain non-overlapping properties, such as the colour of billiard balls, correspond to the negative analogy. The third analogy is the neutral analogy, which refers to properties that are unknown as to whether they are shared. "[T]hese are the interesting properties," Hesse states, because "they allow us to make new predictions" (ibid., p. 8). Of course, the predictions may be incorrect, but that is not a problem. What matters is that they are open to new opportunities in the future due to their unknown nature.

In another context, Ernan McMullin (1976) discusses the "heuristic appraisal" of scientific theory, which also focuses on the unproven future of scientific investigation. In discussing how a theory should be appraised, McMullin argues that instead of focusing solely on the positivist concept of confirmation, the theory's "fertility" should be considered. He distinguishes two types of fertility, each of which requires a different appraisal criterion: "proven fertility" (P-fertility) and "unproven fertility" (U-fertility). While P-fertility confirms the theory's truth-value based on its past success, U-fertility is concerned with the theory's untested future possibilities. The types of appraisal required for each fertility are then identified as epistemic appraisal and heuristic appraisal, respectively.

McMullin's idea is expanded by French (1995, 1997), who claims that the fertility of a theory or model is related to its inherent structure. French's account is noteworthy because it is one of the first philosophical accounts to explicitly discuss model pursuit. This account was built around the question of why, in the late 1960s, physicists pursued the colour model of quarks rather than the paraparticle model, despite a lack of experimental evidence

supporting either model. According to French, physicists pursued the colour model because it provided more opportunities for theoretical development than the paraparticle model. To use McMullin's words, it was because the colour model had higher U-fertility. French labels the degree of opportunities for theoretical development as the "heuristic fruitfulness" of a theory or model, based on Charles Sanders Pierce's concept of "esperable uberty," which refers to the expected fertility of reasoning with little security but will result in new ideas.

French then claims, as a proponent of structural realism, that the colour model was rated as heuristically more fruitful due to its inherent structure. The structure of the colour model, in particular, could be incorporated into a symmetry principle known as gauge invariance, which means that the model could be transformed in specific ways without changing its empirical consequences. Because gauge invariance was incorporated into quantum electrodynamics, dubbed one of the most successful theories in physics, physicists pursued colour models in the hope that gauge invariance would lead to such success.

Another account influenced by McMullin's concept of heuristic appraisal is Šešelja and Straßer's (2014) account on "programmatic character." They, like French, do not clearly distinguish between models and theories and regard models as early-stage theories. According to Šešelja and Straßer, one of the key indicators for a model's (or a theory's) is its programmatic character, which is provided by its heuristic resources. The programmatic character here refers to the character that allows the model to overcome shortcomings such as explanatory anomalies that prevent the model from developing into a successful theory.

Šešelja and Straßer emphasise the model's connection to other theories, similarly to French, as the source of the programmatic character. However, unlike French, the structural similarity is not their focus. Rather, they argue that if the model is embedded in a theoretical framework that addresses the model's shortcomings, the model takes on a programmatic character and can thus be advanced further. The pursuit of Wegner's theory of continental drift serves as a historical example of their claim. According to their illustration, while Wegener's theory could not account for the mechanism of drift in its early stages, the theory of isostasy (a theory about the equilibrium between the crust and the mantle), in which Wegner's theory was embedded, provided a possible mechanism and specified where further research should be conducted.

There are some similarities between French's account of heuristic fruitfulness and Šešelja and Straßer's account of programmatic character that mark those accounts as theory-centric. The first theory-centric feature is about the type of heuristics. Both accounts are concerned with whether the model will develop into a successful

theory. More specifically, French emphasises the model's objective opportunities for future theoretical development, and Šešelja and Straßer emphasise whether the model can overcome its shortcomings and thus grow into a successful theory.

The second theory-centric feature concerns the source of heuristics. Both accounts derive the source from the model's relationship to other theories. French focuses on whether the model shares structural similarities with an already well-developed theory. And Šešelja and Straßer investigate whether the model is embedded in a theoretical framework that can help the model overcome its shortcomings.

This theory-centric understanding of heuristics, however, is not sufficient to capture the heuristic function of models. The pursuit of QSAR models by pharmaceutical scientists in the 1960s is a good example of the limitations of this theory-centric understanding. In the following sections, I will provide an overview of the QSAR model pursuit episode and then revise the concept of the heuristic function of models in light of the episode.

3 A historical overview on QSAR and drug design

Drug design, also known as rational drug design, refers to a set of strategies for developing new drugs that make use of physiological or biochemical knowledge. It began in the mid-twentieth century and has grown in popularity since the 1980s. Some define drug design in terms of "theoretical understanding" in biology and chemistry (e.g., Adam, 2005, p. 514), and it is difficult to deny that some high-level theories, such as the receptor theory of drug action (e.g., Quirk, 2006), played a role in drug design. Nevertheless, focusing solely on high-level theories would mislead us about what actually occurred in the practices of drug design. While high-level theories provided a very basic framework for drug design, they offered little guidance on how to actually design drugs. Instead, knowledge of low-level regularities concerning particular chemical analogues that could be applied at a practical level was crucial for drug design.[7] Specifically, as I shall show in greater detail later, knowledge of the relationship between the chemical structure of specific chemical analogues and their biological activity was

[7] This is supported by an interview with George Hitchings, one of the 1988 Nobel laureates who won the award for his contribution to drug design (Altman, 1988). According to him, the field of chemotherapy in the 1940s was divided between "fundamentalists," who focused on fundamental theories of physiology and biochemistry, and "screeners," who screened a vast number of random compounds. His research group thought that "some kind of middle course might be possible, a course that would generate basic information which chemotherapy could then exploit." In other words, what was essential for his research group was not high-level theories in and of themselves but rather the knowledge that could be applied practically.

critical to drug design.

While the term drug design has not been precisely defined, it has generally been used to emphasise the contrast with the "trial-and-error" approach, which randomly synthesises-and-tests drug candidates until new drugs are discovered.[8] The goal of drug design was to minimise random factors, allowing for the most cost-efficient synthesis-and-testing of chemical compounds. This effort was motivated by the fact that, while there was an abundance of potential drug candidates, chemically synthesising and biologically testing those compounds was hugely expensive until the 1980s.[9] More specifically, while there were nearly an infinite number of chemical compounds possible, the costs of synthesising-and-testing drug candidates were prohibitively high, with one additional synthesis-and-test cycle costing more than 2,000 dollars in the 1960s (Craig et al., 1970, p. 1079). As a result, reducing synthesis-and-test cycles has become a critical issue for pharmaceutical industry scientists to address.

Pharmaceutical scientists used to create catalogues that summarised the chemical structures and biological activities of the compounds they had synthesised-and-tested so far. They hoped that the catalogues would help them decide which compound to synthesise-and-test next. However, the catalogues were unable to organise the synthesis-and-testing data effectively, making it difficult to glean valuable insights from them.

QSAR, invented in 1962–1964 by chemist Corwin Hansch and his colleagues, was a novel method for extracting valuable insights from a collection of data (Hansch & Fujita, 1964; Hansch et al., 1962). QSAR is a computer-based statistical modelling method that predicts the biological activity of a compound based on its molecular structure. Equation 1, shown below, is a QSAR equation presented by Hansch and Fujita (1964, p. 161). The left side of the equation describes the biological activity of a compound using *C*, the concentration of the compound necessary for a particular rate of biological response. The right side describes the chemical structure of the compound using a variety of terms, including π and σ. Here, π is the constant for the lipophilic character (defined as the oil and water partitioning), and σ is the constant for the electronic character (dependent on the electron density at a specific position) of the substituent of the compound.

[8] While the distinction between the "trial-and-error" approach and "rational drug design" is widely recognised, criticisms also exist against this dichotomy (e.g., Lesch, 2008).

[9] Since the 1990s, the development of combinatorial chemistry and high-throughput screening techniques has reduced the cost of synthesis-and-testing.

$$\log\frac{1}{C} = -k\pi^2 + k'\pi + \rho\sigma + k'' \quad (1)$$

However, when QSAR was first developed in the 1960s, it received little attention from pharmaceutical scientists. While working with QSAR models required significant cognitive and material resources, many pharmaceutical scientists did not have enough empirical, theoretical, or even "political" reasons (e.g., academic influence) to justify allocating resources to work with the models.

Let me briefly describe how QSAR models were first developed to highlight the challenges that many pharmaceutical scientists faced when pursuing QSAR models. Hansch's development of QSAR models can be traced back to a debate about plant growth mechanisms. Between the 1940s and the 1960s, Hansch debated with plant physiologist H. Veldstra about the critical factors regulating plant growth (see Veldstra, 1953; Hansch, 1969, 2011). As a physiologist familiar with lipophilic effects, Veldstra asserted that the critical factor is the molecule's distribution between the aqueous and fatty phases of a cell. This claim was consistent with the intuition of many pharmaceutical scientists that the lipophilicity of molecules relates to their biological activities. In contrast, Hansch's early work (1940s–1950s) concentrated on physical organic chemistry, and thus on electron density at specific positions of the molecule.

What enabled the development of QSAR was Hansch's decision to concentrate on both lipophilic *and* electronic effects concurrently. This decision was made possible because Hansch was one of the few chemists in the 1960s who had access to a computer for research purposes.[10] Before the widespread adoption of computers, scientists sought statistical relationships between only two variables, since it was nearly impossible to correlate more than two variables simultaneously using only pen-and-paper. In comparison, as one of the early computer users in his field, Hansch was able to investigate more than two variables. The result of the multiple parameter analysis was the first QSAR model, which statistically correlates the three variables—lipophilic, electronic, and biological—in one equation.

However, those characteristics of QASR that made it novel also served as obstacles for pharmaceutical scientists. First, the incorporation of the concepts from physical chemistry strengthened the pharmaceutical industry chemists' opposition to QSAR. The "extrathermodynamic" approach was another name for QSAR. That

[10] Hansch could use a computer donated to the Chemistry Department by a College trustee in 1961 (Hansch, 2011, p. 502).

was because QSAR parameters were characterised using thermodynamic concepts ("extra-*thermodynamic*"), whereas the relationships described between structure and activity were not included in the formal structure of thermodynamics ("*extra*-thermodynamic"). In comparison, the majority of chemists working for pharmaceutical companies were trained in synthetic organic chemistry. The organic chemists were perplexed by QSAR's numerical molecular description based on physical terms.

Along with the unfamiliarity with physical chemistry, computers and multiple parameter analysis were the most significant barriers to pharmaceutical scientists' use of QSAR. By the 1960s, the majority of pharmaceutical scientists lacked experience with computer-assisted research.[11] Although computers were introduced to pharmaceutical companies before the 1960s, they were initially used primarily for accounting purposes rather than scientific research. Even in some pharmaceutical companies that had early adopted computers for research purposes, computer usage was severely restricted, as all personnel from various departments were required to share a very limited number of computers (see, e.g., Craig, 1971a, p. 160; Boyd, 2007, p. 403).

Nevertheless, *some* pharmaceutical scientists in the 1960s believed that QSAR could aid in the development of new drugs. Paul N. Craig, a chemist who worked for Smith Kline and French (SK&F; now GlaxoSmithKline) and whom I will discuss in greater detail in Section 6, was one of the pharmaceutical scientists. What drew the scientists' attention mostly was QSAR's potential for use as a guide for synthesis-and-testing. In scientific articles on QSAR published in the 1960s and 1970s, pharmaceutical scientists frequently expressed their hope that QSAR models would reduce the number of compounds required for synthesis-and-testing. For instance, Craig and his colleagues stated:

> "During the chemical synthetic and biological testing programs....the original publications of Hansch and coworkers [in the 1960s] came to our attention. We recognised that a prompt application of Hansch's correlative techniques, if successful, promised to give valuable guidance to our synthetic [and testing] efforts, and to reduce the number of compounds which we would need to synthesize and test." (Craig et al., 1970, p. 1079)

In the 1960s, with Craig as the primary link, SK&F funded Hansch's research and collaborated with him.[12]

[11] For a history of computer use in the pharmaceutical industry, see, e.g., Gambardella (1995), Brooks and Gmyrek (2011); for a history of computer use in (computational) chemistry, see., e.g., Gavroglu & Simões (2012).

[12] Hansch acknowledges SK&F for their financial help (Hansch, 1969, p. 239).

Following SK&F, several pharmaceutical companies began supporting and/or collaborating with Hansch.[13] Based on the pioneering efforts of the pharmaceutical scientists, experimental data supporting QSAR models began to be collected. In the 1980s, the term "drug design" became very popular as an alternative to the trial-and-error approach, and many pharmaceutical scientists saw QSAR as one of the most important methods for driving drug design.[14]

The historical question here is why some scientists in the 1960s found QSAR attractive. More specifically, what aspects of QSAR models led pioneer scientists like Craig at SK&F to believe that QSAR would reduce the number of compounds required for synthesis-and-testing? While the existing literature on model pursuit emphasises the model's connection to theory, the pharmaceutical scientists' belief was not derived from QSAR's connection to theory; as I previously described, QSAR was also known as the extrathermodynamic approach due to its significant deviation from thermodynamics or any other theories. It was also not due to empirical evidence; because few people worked on QSAR models until the 1960s, there were few model outputs and experimental data to compare. Even Hansch's prominence or academic influence was ineffective; Hansch was a chemistry professor at Pomona College, a small liberal arts college. His background, early academic performances, and the memoir he wrote show that Hansch was not well-known at least until the 1960s.[15]

To comprehend this historical episode, it is necessary to revisit the concept of the heuristic function of models. In Section 4, I will revise the concept of heuristic function in light of this episode. Then, in Sections 5 and 6, I will go into greater detail about how models perform the heuristic function.

4 The heuristic function of models revisited

4.1 Models as epistemic tools

In Section 2, I reviewed the existing literature on the model pursuit and identified its theory-centric features.

[13] Hansch acknowledges Eli Lilly Company (Hansch & Anderson, 1967, p. 753), Hoffman-La Roche, Inc., and Chas. Pfizer & Co., Inc. (Hansch et al., 1968, p. 11) for their support of the samples of chemical compounds.

[14] For example, at the Third Rhone-Poulenc Round Table in November 1982, scientists identified QSAR as their most preferred optimization technique for drug design (Jolles & Wooldridge, 1984, pp. 242-244).

[15] Some stories in Hansch's memoir reflect his poor reputation at the time. For example, fearing that no good scientist in the United States would want to collaborate with him at the small liberal arts college, he focused on finding a foreign postdoctoral associate willing to visit the country (Hansch, 2011, p. 502). Furthermore, some pharmaceutical and pesticide research directors mocked his QSAR research (ibid., p. 497).

The first feature was the focus on the theoretical development of the models; models have heuristic potential, according to the existing literature on model pursuit, in the sense that they may develop into successful theories in the future. This approach to heuristics overlaps with traditional views on models that regard models as adjuncts to theories.

However, since the rise of Morgan and Morrison's (1999) view on models as mediators, model literature has widely acknowledged that models can function as investigative instruments that are partially independent of both theories and the world. Among the recently developed views that are influenced by Morgan and Morrison, a view of models as "epistemic tools" emphasises that models are intentionally constructed artefacts for a variety of tasks (e.g., Knuuttila, 2011; Knuuttila & Boon, 2011; Currie, 2017). The epistemic tool view distinguishes itself from other views of models in that, rather than focusing solely on the representative relationship between models and their target systems, it stresses that models are embodied in the (material) medium that model users can manipulate. In other words, this view perceives models not as merely abstract structures, but rather as tools in the real world that are purposefully constructed in a medium to accomplish a variety of specific tasks.

This view of models as epistemic tools provides a better understanding of the practices scientists engage in when pursuing models. While existing pursuit literature asks whether models will develop into successful theories about their targets, models themselves do not have to be considered potential theories if we consider models to be tools. Rather, models can serve their heuristic function by acting as tools for model users, performing various specific tasks that assist model users in efficiently obtaining new knowledge in the future. This knowledge may not be of the kind that is used to develop new theories in the near future or is derived from theories. However, it assists model users in organising and directing their research practices and achieving a variety of specific goals.

This type of heuristic function of models has been examined multiple times outside of the literature on pursuits. For instance, Miles MacLeod's (2015) study on the heuristic role of models in systems biology shows how models can perform this type of heuristic function. He demonstrates how models can assist scientists in addressing the issue of collaboration between modellers and experimenters by assisting in the generation of hypotheses that direct experiments. This heuristic function differs from what is described in the existing literature on model pursuit. Although both focus on the future advancement of scientific research rather than proven success, MacLeod's discussion does not see the models as the ones that should grow into successful theories. Rather, he sees models as tools that direct the practices of model users in particular contexts.

Yafeng Shan (2020) provides another comparable account to the epistemic tool view. He evaluates the "promisingness" of a theory by treating the practices provided by the theory as "tools" that can be used for specific purposes. Yet, a crucial difference between Shan's account and what I claim in this paper by treating models as epistemic tools is that, whereas Shan exclusively focuses on whether a theory will provide useful practices *after* its further development, I also focus on the process *during* the pursuit of a model. According to Shan's view, when evaluating the promisingness of a theory, we should focus on the problem-defining and problem-solving power that the theory could potentially contribute after its development is complete. My approach, on the other hand, does not wait until the model's development is complete. I am not saying here that the model's guidance is not forward-looking; it is, in the sense that it is aimed at achieving epistemic goals in the future. Yet, it is *through* the process of pursuing, i.e., working with the models, that the models play their heuristic function and model users move closer to their epistemic goals.

The pursuit of QSAR models can be better explained if we centre our attention on this type of heuristic function. The reason pharmaceutical scientists pursued QSAR models in the 1960s was not because they expected them to be successful theories. Instead, they expected QSAR models to provide "valuable guidance" to their "synthetic [and testing] efforts" in the face of extremely high synthesis-and-testing costs (Craig et al., 1970, p. 1079). Thus, the point that needs to be articulated to capture the pursuit of QSAR models is not about the models' theoretical development, but about how the models guided the "synthetic [and testing] efforts." And as I will show below, this role of models was linked to how models assist model users in assessing specific conditional hypotheses.

4.2 The testworthiness of hypotheses

If models perform their heuristic function not as potential theories but as epistemic tools, we would need to clarify how the models assist in knowledge acquisition without relying on the development of theories. To specify this kind of heuristic function, I suggest that we shift our focus from theories to hypotheses. To be clear, by "hypotheses" in this paper, I do not mean high-level theories. Rather, I mean very focused conditional hypotheses about low-level regularities, which is similar to what Ian Hacking (1983) means by low-level generalisations. By distinguishing low-level generalisations from high-level theories, Hacking highlights the philosophical importance of the manipulation of entities, rather than abstract theories. In this section, I extend the discussion on

low-level regularities towards the topic of model pursuit by raising an issue regarding the costs required to manipulate the entities.

When I separate low-level regularities from high-level theories, I do not imply that they are clearly separated. Rather, what I want to highlight is that knowledge concerning low-level regularities has some characteristics that are difficult to find in high-level theories. First, they are concrete and flexible. QSAR models, for example, were concerned with the low-level regularities of a series of chemical analogues (e.g., Craig et al. (1970) for the 3-tropanyl 2,3-diarylacrylates) rather than a general theory that unifies the properties of the whole chemical and biological world. Since QSAR models were concerned with low-level regularities, a QSAR model based on a particular series of analogues could not guarantee that it would function well for other types of compounds with significantly different structures. Another related characteristic is that experimental manipulation often plays a central role in the knowledge of low-level regularities. In other words, the questions about the manipulation of specific entities at hand take precedence over other questions, such as whether the knowledge is consistent with other theories. These characteristics of low-level regularities are frequently referred to as chemistry's characteristics (e.g., Simon, 2012; Schummer, 2021). But there is no need to limit the scope to chemistry alone.

As stated in Section 3, drug design was developed as an alternative to the trial-and-error approach to reduce the randomness inherent in drug discovery. Here, it was critical to find and formalise the regularities relating to the chemical and biological properties of drug candidates, rather than leaving them to serendipity. During the early stages of drug design, each drug candidate was linked to particular conditional hypotheses, which typically took the form "if compound *C* has structure *S*, then it shows biological activity *A*." The challenge was that learning whether the hypotheses were true was exceedingly expensive because the number of possible hypotheses was nearly infinite (as was the number of possible structures), and observing biological activity required huge costs. In drug development in the real world, where time and resources are limited, the high expense made it difficult, or often impossible, to develop a new drug. Thus, the success of drug design depended on identifying which of the conditional hypotheses was more worthwhile testing, i.e., worthy of being allocated resources to be tested.

To emphasise this point, I propose the term "testworthiness" of hypotheses.[16] While the logical and abstract

[16] To clarify, while the testworthiness and the pursuitworthiness of hypotheses are related, they are not identical. Testworthiness is a sub-concept of pursuitworthiness that focuses on testing practices, among other pursuit-related practices.

aspects of testing have been highlighted in the traditional philosophy of science, the costs associated with testing have frequently been ignored. For example, Popperian falsificationism treats all hypotheses that have not been explicitly refuted or replaced as remaining on the table (Popper, 1959). Thomas Nickles (2006), however, criticises this eliminative methodological strategy by raising concerns about the costs associated with testing the hypotheses. Given the costs associated with testing, scientists must determine whether their hypotheses are worth testing. The majority of untested hypotheses are kicked-off the table before they can be explicitly tested. Nickles says, "[s]cientific claims are in effect sterile unless and until someone finds them worth testing or at least worth citing" (ibid., p. 164). If we take Nickles' claim seriously, we can say that the purpose of identifying testworthy hypotheses is to provide opportunities for their release from sterility.

The identification of the testworthiness of hypotheses is the point at which models can contribute to the efficient acquisition of knowledge without relying on theoretical development. In other words, models can play their heuristic function by assisting model users in identifying the testworthiness of hypotheses. Sometimes, of course, scientists may want their models to grow into successful theories. In everyday scientific practice, however, scientists frequently attempt to deal with a variety of highly specific hypotheses that they are now confronted with, rather than directly attempting to develop high-level theories. A detailed explanation of the method by which models assist in identifying the testworthiness of hypotheses will be provided in Sections 5 and 6. But before moving on to Sections 5 and 6, let me briefly introduce below the core of the method.

4.3 Material and technological medium of models

In discussing the heuristic function of models thus far, I have switched the focus to where models perform a heuristic function related to the testworthiness of low-level hypotheses. Now I would like to address another, but related, implication of the epistemic tool view: models are not merely abstract structures, but rather are embodied in the material medium that allows model users to interact with them. Based on the epistemic tool view's emphasis on the medium of models, I propose that the heuristic function of models be conceived as resulting from the interaction of the model users with the models' medium. By interaction, I mean not only that model users influence the models by building or manipulating them but also that the models influence the model users. In particular, I emphasise that models enable model users to recognise phenomena in specific ways, which will be the main theme of Section 5.

When explaining what motivates the heuristic function of models, the existing literature on model pursuit focuses on the abstract and theoretical contents of models. For example, as stated in Section 2, French's "heuristic fruitfulness" and Šešelja and Straßer's "programmatic character" highlight the model's connection to another existing theory. However, this theory-centric approach cannot capture the pursuit of QSAR models. As the term "extra-thermodynamic approach" implies, QSAR models deviated significantly from thermodynamics or any other theories.

If not theories, what did motivate the heuristic function of QSAR models? Both French's and Šešelja and Straßer's accounts that highlight the connection to existing theory, I believe, are derived, at least in part, from their assumption that the goal of scientists pursuing a model is to develop it into a successful theory. This theory-centric assumption pushes us to seek a hint or support from an existing successful theory in developing the model into another successful theory. However, as I previously argued, models can perform the heuristic function by serving as guiding tools rather than growing into successful theories on their own. The point here is that, as highlighted in the view of models as epistemic tools, the goal of using tools varies depending on the specific research contexts; tools can perform a variety of tasks depending on the research contexts, and even a tool that was deemed useful in one context may be ignored in another. And, if the goal of the heuristic function is not fixed but varies, why don't we recognise that what motivates the heuristic function varies as well?

The view of models as epistemic tools emphasises that models function in diverse ways depending on the material medium in which they are embodied (e.g., Knuuttila, 2005; Knuuttila & Boon, 2011). For example, a model inscribed on paper with diagrams and another model displayed on a computer screen with 3D graphics would be built and manipulated differently, promote different types of scientific reasoning, and serve different purposes. The heuristic function of models, I believe, cannot be an exception. To comprehend what motivates the heuristic function of models, we should consider not only their abstract contents but also the material and technological medium in which they are embodied and how model users interact with that medium. And to discuss the pursuit of models, we should consider whether the costs required to interact with the material and technological mediums would be worthwhile.

In this vein, the medium of QSAR models should be considered when discussing their heuristic function. As computational models utilised for specific chemical and biological purposes, QSAR models possessed notable material and technological characteristics. Among the various aspects of QSAR models, what I focus on in this

paper is that QSAR models were one of the first computational models in pharmaceutical research to be embodied with computers capable of doing multiple parameter regression analysis. Such material and technological mediums endowed QSAR models with specific ways of functioning, which aided pharmaceutical scientists in the study and development of new drugs. Specifically, by interacting with the medium of QSAR models, pharmaceutical scientists could recognise certain phenomena concerning the chemical and biological properties of drug candidates. The next section will go into greater detail on the recognition of phenomena, or what I will term the "construction" of phenomena.

5 The construction of phenomena

Now that we have a revised concept of the heuristic function of models, the question is *how* models perform this heuristic function. While existing pursuit literatures focus on the connection of models to theories, I focus on the relationships of models with data and phenomena. I argue, in particular, that models fulfil their heuristic function by "constructing" phenomena from data, allowing model users to recognise the phenomena *as such*. In this section, I discuss the construction of model phenomena. Further defence of how the constructed phenomena are related to the heuristic function of models will be provided in Section 6.

Looking back through the history of philosophical discussions on models, it has been typically assumed that models have already specified target phenomena. However, several different approaches have recently raised some challenges to the assumption. Among them, the one on models' "exploratory" role is particularly relevant to the context of pursuit. The exploratory role refers to the early investigation of scientific phenomena or problems in the absence of fully formed or readily applicable underlying theories (e.g., Gelfert, 2016; Fisher et al., 2021). In a discussion of the exploratory function of models, Axel Gelfert (2018) criticises traditional views for assuming that models are always intended for already specified target phenomena. He then claims that models can be used *in search of* target phenomena.[17]

The epistemic tool view of models has also addressed a similar point. According to Knuuttila and Boon (2011),

[17] Gelfert claims that there are several distinct aspects of exploratory models that can contribute to the search for target phenomena. They include: as "starting points for further inquiry," providing "proofs of principle," providing "potential explanations," and the "search for" potential target (Gelfert, 2018, pp. 9-12). It should be noted, however, that this does not imply that an exploratory model should not have target phenomena. Cope and Hardy's exploratory models, for example, provided possible explanations for two molecular rearrangements called the Claisen and Cope rearrangements (Fisher, 2006; Gelfert, 2016, pp. 87-93).

the current analysis of the representation of scientific models is overly focused on ready-made models. In opposition to the prevalent representational approach, they advocate paying attention to the modelling process because they believe that the representational relation between models and their target phenomena is not pre-determined but is established throughout the modelling process. Importantly, what they regard as being established during the modelling process rather than pre-determined does not only include models, but also the target phenomena; they claim that the target phenomena are "co-constructed along with the models" (ibid., p. 319). Thus, while they do not expressly limit their focus to exploratory research, we can say that their argument is aligned with the discussion on exploratory models in that they emphasise that the target phenomena of models are not necessarily pre-specified but are "constructed" through the construction and manipulation of models.

But what does it mean to say that the target phenomena are constructed? While Knuuttila and Boon develop their account using the term "construction" of target phenomena, they do not clarify what they mean by the term. Here, by using Daniela Bailer-Jones' (2009, ch. 7) discussion of the role of models in phenomena recognition, I want to clarify the term: models construct their target phenomena in the sense that they enable model users to recognise phenomena in particular ways.

Bailer-Jones develops her account based on Bogen and Woodward's (1988) account of the distinction between data and phenomena. According to Bogen and Woodward, phenomena are rarely observable in and of themselves but are inferred from observable data. They go on to say that inferring phenomena from data is not "theory-laden," because data are not explained by theories. Woodward (2011) clarifies that their claim does not mean theories are not involved; theories can be involved in many ways, such as by providing a vocabulary for characterising data and phenomena. Rather, he contends that theories do not always serve an explanatory role that allows data to be used as evidence for phenomena.

Building on this data-phenomena distinction, Bailer-Jones emphasises the role of models in inferring from data to phenomena. Phenomena are more than just patterns in data sets; they require a specific description to be identified as phenomena. And, according to Bailer-Jones, models can provide such descriptions. In other words, models can help model users recognise previously unidentified phenomena. Bailer-Jones's idea is comparable, but not identical, to Ian Hacking's (1983, ch. 13) concept of "creation of phenomena." Using this term, Hacking emphasises how extremely unusual experimental conditions enable scientists to observe phenomena that never or rarely occur outside the laboratory (e.g., Hall effect), in contrast to the conventional view of phenomena, which

assumes they occur naturally. Bailer-Jones, on the other hand, is more concerned with our recognition of the phenomena, regardless of how frequently the phenomena occur in nature.

Borrowing from Bailer-Jones's discussion, I argue that models construct phenomena from data in the sense that they allow model users to recognise phenomena in specific ways. As many philosophers might believe, phenomena may exist in nature regardless of our recognition. They cannot, however, be identified as phenomena unless we recognise them *as such*.

The concept of phenomena construction is critical for comprehending QSAR's heuristic function. As stated in Section 3, QSAR models were characterised by the incorporation of multiple parameters. Specifically, one of the most novel aspects of QSAR was that it "put the hydrophobic [or lipophilic] effect… on a scale similar to… electronic effects" (Martin, 2012, p. 61). Before QSAR, many pharmaceutical scientists intuitively knew that lipophilicity correlates with biological activity, and some other scientists recognised the critical role of electronic effects in biological activity. Nonetheless, they were unaware that lipophilic and electronic effects are simultaneously linked to biological activity *in concert*, on a similar scale. But by constructing and manipulating QSAR models embodied in computers capable of regression analysis with multiple parameters, pharmaceutical scientists could recognise the correlation between lipophilic, electronic, and biological characteristics of chemical compounds. It was the correlation that the scientists "would not have found" in the absence of "the systematic approach developed by Hansch" (Craig et al., 1970, p. 1079). Thus, in the sense that QSAR enabled its users to recognise phenomena as such, I contend that QSAR "constructed" phenomena.

In the next section, I will explain how the construction of phenomena results in the efficient acquisition of knowledge and the achievement of long-term epistemic goals. The point will be that it is not because the models accurately represent the world or because (the description of) the constructed phenomenon is accurate. Instead, I will show that by constructing phenomena, models can guide the research practices of model users by aiding in the identification of the testworthiness of hypotheses.

6 From phenomena construction to hypotheses testworthiness: The case of the Craig plot

This section elaborates on how the models' construction of phenomena contributes to the models' heuristic function, using the case of Paul N. Craig's pursuit of QSAR models. As I mentioned in Section 3, Craig, a medicinal chemist at SK&F, was an early user of QSAR models. In this section, I show that the constructed

phenomena enabled scientists to see a chemical compound not only as a potential drug but also as a piece of a puzzle that aids in the future development of a drug. This change was possible because constructed phenomena provided criteria for distinguishing hypotheses whose testing enables new and valuable learning from those whose testing provides little new information. And using the criteria, the model users were able to determine which hypotheses were more testworthy, i.e., more deserving of investing resources in testing practices.

The main point of the concept of testworthiness of hypotheses, which is a concept I proposed in Section 4, is that because hypothesis testing incurs costs, we must carefully consider whether the hypothesis is worth testing at those costs. For example, in drug design, in order to test the hypothesis that "if compound *C* has structure *S*, then it shows biological activity *A*," data on compound *C*'s biological activities *A* must be produced. The data production begins with the synthesis of compound *C* and continues with its introduction into specific experimental settings that capture certain aspects of biological systems. Following that, the produced data must be properly formatted and handled before being compared to the predicted results of the hypothesis. Given the extreme complexity of biological systems, these processes typically entail significant material and intellectual costs. When we say a hypothesis is testworthy, we mean that it is worthwhile to produce and compare the associated data given those costs.

In actual scientific practices, especially in the early stages of research, scientists face the challenge of determining which hypotheses are more testworthy than others. Specifically, at the beginning stages of research, when there are too many possibilities open and too many hypotheses available, which hypotheses should receive the resources for testing? There would be a variety of conditions that make a hypothesis testworthy. Regarding the issue of resource allocation, however, I would want to emphasise that we should have *low* confidence in the results of hypothesis testing. Otherwise, if we are sure that the hypothesis's testing results will match the predicted results (i.e., prediction about the empirical consequences) that we had in mind prior to testing the hypothesis, then investing in the costs of testing the hypothesis would be pointless. To be clear, I do not assert that the hypothesis with the lowest level of confidence is always the most testworthy. The point is rather that we should lack confidence to some extent and be able to distinguish between hypotheses with lower and higher levels of confidence. And if all things being equal, the lack of confidence makes the hypothesis more testworthy than its competitors.[18]

[18] The fact that scientists often test hypotheses in which they have a high level of confidence might appear to

This idea about confidence in the testing results is echoed in Rune Nyrup's (2015) account of inference to the best explanation. Many people believe that a hypothesis is a better explanation than its rival if it is more likely to be true. Nyrup, on the other hand, proposes an alternative criterion for a good explanation that does not rely on whether a hypothesis is likely to be true; he argues that a hypothesis can be a better explanation than its rival if learning whether it is true is more epistemically valuable. The epistemic value here is determined by the agent's epistemic goals. More specifically, as the results of hypothesis testing (e.g., data for/against the hypothesis) become more effective at achieving epistemic goals, the hypothesis can more effectively help us to increase our understanding of the phenomena, and thus the hypothesis becomes more epistemically valuable. Using Nyrup's term, we can say that if we have a high level of confidence that a hypothesis's test results will match the previously predicted results, the hypothesis is not highly epistemically valuable because there is little we can learn from testing it.

The point at which models perform their heuristic function by constructing phenomena is *during* the distinction of hypotheses according to their levels of confidence. Specifically, I argue that by constructing phenomena, models provide a criterion for the model users to assess how confident they are about the results from testing the hypothesis. And by distinguishing the hypotheses with a lower level of confidence from those with a higher level of confidence, the model users can identify which hypotheses are more testworthy.

Before the introduction of QSAR models, despite the importance of identifying more testworthy hypotheses about candidate drugs, it was nearly impossible to make such an identification. The test results for nearly all possible hypotheses were highly and equally uncertain due to the extreme complexity of chemical and biological systems. Thus, the typical approach at the time was the "trial-and-error" approach, which was to keep synthesising-and-testing the candidate drugs, relying heavily on random factors, until the compound showed some

---

refute my claim here. However, we should note that this type of testing typically occurs in the later stages of research, such as when the chemical properties of compounds that a scientist desires are highly specified due to previous testing of their analogues. In this case, because the scientist has already specified the desired properties of compounds to some extent, the scope of interests may be limited to a certain range of hypotheses with high confidence. Yet, even if the scientist has high confidence in these hypotheses, it is necessary for them to have some degree of uncertainty in order to be testworthy; if the scientist already knows the testing results for certain prior to the actual testing, they cannot learn anything new from the testing. More specifically, when there are two hypotheses *A* and *B* concerning each corresponding chemical compound, it is not that *A* is more testworthy than *B* because the researcher has a higher confidence in *A* than in *B*. Rather than that, a more plausible explanation is that only *A* is considered since only *A*, and not *B*, fits the scope that the researcher specified, and *A* is testworthy because the scientist still lacks confidence in *A* to some extent. But if *A* and *B* all fit the scope of the researcher and all things being equal, the lack of confidence would make *B* more testworthy than *A*.

desired biological activities.

This challenge associated with the identification of testworthy hypotheses explains why some pharmaceutical scientists found QSAR appealing. When Craig and his colleagues at SK&F collaborated with Hansch in the 1960s, they initially synthesised-and-tested only nine compounds (Craig et al., 1970). They then gave Hansch the data from the synthesis-and-testing, and Hansch constructed a QSAR model using the data. This QSAR model showed a correlation between lipophilic, electronic, and biological characters, i.e., it constructed the phenomena that they did not recognise before. The next compounds to test-and-synthesise were chosen based on the correlation.

Craig (1971b) went into greater detail about this type of process. Using some data obtained from previously synthesised-and-tested candidate drugs, Craig generated graphs (dubbed "Craig plot") with lipophilic (π) and electronic (σ) axes and drew dots to indicate the data from corresponding candidate drugs (see Fig. 1). This graph visualised the correlation between lipophilic, electronic, and biological characters. Using this graph, Craig was able to determine which candidate drugs are more or less worth synthesising-and-testing. For example, one should avoid synthesising-and-testing chemical compounds that "lie on or near" already synthesised-and-tested compounds on the graph (ibid., p. 684). Due to the proximity in the graph, it was possible to predict with high confidence the results of synthesising-and-testing the compounds. In contrast, if a compound was placed far from other compounds on the graph, it was difficult to predict the results of its synthesising-and-testing.

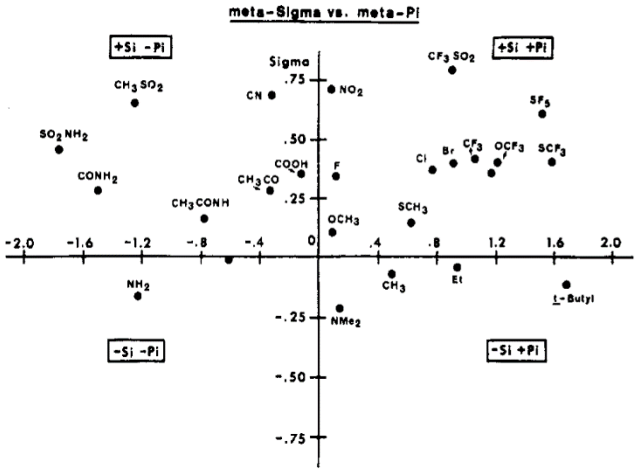

Fig. 1 Craig plot (reprinted with permission from Craig (1971b, p. 683). Copyright (1971) American Chemical Society)

Fig. 2 is a hypothetical graph that I drew to make the Craig plot easier to comprehend. Let the black circles be the data from the hypotheses that have already been tested ($H_{\text{black circle}}^{(1)}$, $H_{\text{black circle}}^{(2)}$,…, $H_{\text{black circle}}^{(n)}$), and let

the red square and blue triangle be the predicted results for the hypotheses that have not been tested ($H_{red\ square}$ and $H_{blue\ triangle}$, respectively). Scientists can easily guess the testing results of $H_{blue\ triangle}$ with high confidence because the blue triangle is located very close to black circles, i.e., data obtained from previously tested hypotheses $H_{black\ circle}^{(1)}$, $H_{black\ circle}^{(2)}$,…, $H_{black\ circle}^{(n)}$. In this case, testing whether $H_{blue\ triangle}$ is true is less valuable to the scientists because they can know the results almost definitively even before actually testing it. By contrast, the truth-value of $H_{red\ square}$ is highly uncertain because the red square is located far from the black circles. In this case, learning about $H_{red\ square}$ through testing becomes valuable, as the test results would provide new and valuable information.

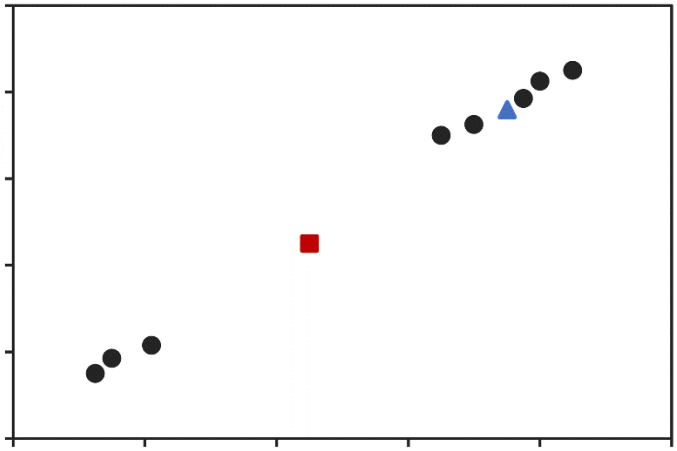

Fig. 2 A hypothetical graph drawn by the author. Black circles are the data for hypotheses that have already been tested, and the red square and blue triangle are predicted results for hypotheses that haven't been tested yet

Technically, the words such as "close" and "far" that I use to describe Fig. 2 are not precise enough. There are many statistical techniques for defining and measuring the distance between hypotheses.[19] However, the point I want to make here is that the uncertainty associated with hypotheses, i.e., the level of confidence in whether the predicted results will match the data obtained from testing hypotheses, is estimated using their relationship to the phenomena that models construct. And it is through this process that scientists can efficiently identify the testworthiness of hypotheses.

[19] In a regression analysis, as a hypothesis becomes more "far" from the other hypotheses used to estimate the regression coefficient, the prediction about the hypothesis becomes more uncertain, because the prediction will exceed the confidence interval; Yvonne C. Martin, a medicinal chemist, describes the distance between hypotheses as the separation of physical properties between drug candidates. She states, "[d]epending on the difficulty of synthesis and testing one should consider including enough analogs that there would be a good separation of the relevant physical properties." (Martin, 1978., p. 268)

This idea of measuring the level of confidence in a hypothesis based on the hypothesis's link to constructed phenomena is in line with Hesse's (1970) tripartite distinction and French's (1995, 1997) heuristic fruitfulness, which I discussed earlier in this paper. While highlighting the heuristics of models for the unknown future, Hesse and French also point out that the heuristics should be grounded in certain systematic links with some known theories or hypotheses. In particular, Hesse's discussion of neutral analogies is grounded on the tripartite distinction of analogies that include positive and negative analogies. And French claims that the structural correspondence between some of the elements of theories or models establishes the initial bridge between them and thereby contributes to the heuristic fruitfulness of the models. In a similar manner, I am suggesting here that phenomena are constructed upon previously tested hypotheses, and the level of confidence in an untested hypothesis is evaluated based on its relationship to the constructed phenomena.

In a paper about the effects of QSAR models on the pharmaceutical industry, a scientist drew a distinction between views on a chemical compound as "a potential candidate for development" and as "a piece in a puzzle providing information that may eventually lead to an optimal compound" (Unger, 1980, p. 50). By using QSAR models, or more specifically, by recognising the phenomena that were constructed by the QSAR models, pharmaceutical scientists were able to treat each chemical compound as a piece of a puzzle for the future rather than as an isolated entity. If we only consider whether QSAR models accurately described phenomena or had connections to other theories, the pursuit of QSAR models will remain a mystery. Rather, we should acknowledge that the value of QSAR models resided in the fact that they provided new lenses through which to treat and evaluate chemical compounds.

7 Concluding remarks

To sum up, I challenged the current theory-centric approach to model pursuit by developing an account of model pursuit based on the view of models as epistemic tools. The point was that rather than considering models themselves as potential high-level theories, we should be asking if investing in costs to interact with the models will contribute to the efficient acquisition of knowledge about low-level regularities. Models construct phenomena from data; although models may not create phenomena in the ontological sense, they do construct phenomena in the sense that they allow their users to recognise the phenomena as such. And it is through the construction of phenomena that models fulfil their heuristic function, without necessarily growing into or relying on high-level

theories. Models, by constructing phenomena, can aid in identifying what I call the "testworthiness" of hypotheses and thus efficiently guide our epistemic practices in the material world.

I believe that the attitude of "taking model pursuit seriously" can aid both practitioners and philosophers of science today. For instance, the recent rise of machine learning models for drug development is comparable to the rise of QSAR models for drug design decades ago. As this paper has shown, the pursuitworthiness of QSAR models lies not exclusively in their ability to accurately represent the world but also in their ability to guide research efforts at the intersection of chemistry and biology. Such a philosophical investigation of the historical episode of QSAR model pursuit could help pharmaceutical scientists figure out how to choose machine learning models they will work with. Even though it will not be a definitive answer for scientists, the philosophical insights will serve as "pursuitworthy" guidelines, encouraging interdisciplinary dialogue between philosophers and scientists.

References

Adam, M. (2005). Integrating research and development: the emergence of rational drug design in the pharmaceutical industry. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *36*(3), 513-537. https://doi.org/10.1016/j.shpsc.2005.07.003

Altman, L. K. (1988, October 18). 3 Drug Pioneers Win Nobel in Medicine. *The New York Times*. https://www.nytimes.com/1988/10/18/science/3-drug-pioneers-win-nobel-in-medicine.html

Ariëns, E. J. (1971). A general introduction to the field of drug design. In E. J. Ariëns (Ed.), *Drug Design* (Vol. 1) (pp. 1-270). New York: Academic Press.

Bailer-Jones, D. M. (2009). *Scientific models in philosophy of science*. Pittsburgh: University of Pittsburgh Press.

Bogen, J., & Woodward, J. (1988). Saving the phenomena. *Philosophical Review*, *97*(3), 303-352. https://doi.org/10.2307/2185445

Boyd, D. B. (2007). How computational chemistry became important in the pharmaceutical industry. In K. B. Lipkowitz, & T. R. Cundari (Eds.), *Reviews in Computational Chemistry* (Vol. 23) (pp. 401-451). London: John Wiley & Sons.

Brooks, D. A., & Gmyrek, D. P. (2011). Computers come into pharmacy practice: the first retail pharmacy computer system in the state of Virginia. *Pharmacy in History*, *53*(2/3), 113-115.

Craig, P. N. (1971a). Comparison of batch and time-sharing computer runs for correlating structures and bioactivity by the Hansch method. *Journal of Chemical Documentation*, *11*(3), 160-162. https://doi.org/10.1021/c160042a009

Craig, P. N. (1971b). Interdependence between physical parameters and selection of substituent groups for correlation studies. *Journal of medicinal chemistry*, *14*(8), 680-684. https://doi.org/10.1021/jm00290a004

Craig, P. N., Caldwell, H. C., & Groves, W. G. (1970). Spasmolytics. 3. 3-Tropanyl 2, 3-diarylacrylates. A case history of pi-sigma structure-function correlation. *Journal of medicinal chemistry*, *13*(6), 1079-1081. https://doi.org/10.1021/jm00300a014

Currie, A. (2017). From models-as-fictions to models-as-tools. *Ergo: An Open Access Journal of Philosophy*, *4*(27). 759-781. https://doi.org/10.3998/ergo.12405314.0004.027

DiMarco, M., & Khalifa, K. (2022). Sins of inquiry: How to criticize scientific pursuits. *Studies in History and Philosophy of Science*, *92*, 86-96. https://doi.org/10.1016/j.shpsa.2021.12.008

Fisher, G. (2006). The autonomy of models and explanation: anomalous molecular rearrangements in early twentieth-century physical organic chemistry. *Studies in History and Philosophy of Science Part A*, *37*(4), 562-584. https://doi.org/10.1016/j.shpsa.2006.09.009

Fisher, G., Gelfert, A., & Steinle, F. (2021). Exploratory models and exploratory modeling in science: Introduction. *Perspectives on Science*, *29*(4), 355-358. https://doi.org/10.1162/posc_e_00374

French, S. (1995). The esperable uberty of quantum chromodynamics. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, *26*(1), 87-105. https://doi.org/10.1016/1355-2198(95)00006-f

French, S. (1997). Partiality, pursuit and practice. In M. L. Dalla Chiara, K. Doets, D. Mundici, & J. van Benthem (Eds.), *Structures and Norms in Science: Proceedings of the 10th International Congress on Logic, Methodology and Philosophy of Science, 19-25 August 1995* (pp. 35-52). Dordrecht: Kluwer.

Gambardella, A. (1995). *Science and innovation: The US pharmaceutical industry during the 1980s*. New York: Cambridge University Press.

Gavroglu, K., & Simões, A. (2011). *Neither physics nor chemistry: A history of quantum chemistry*. Cambridge:

MIT Press.

Gelfert, A. (2016). *How to Do Science with Models: A Philosophical Primer*. Dordrecht: Springer.

Gelfert, A. (2018). Models in search of targets: Exploratory modelling and the case of Turing patterns. In A. Christian, D. Hommen, N. Retzlaff, & G. Schurz (Eds.), *Philosophy of Science: Between Natural Sciences, Social Sciences, and Humanities* (pp. 245-271). Dordrecht: Springer.

Hacking, I. (1983). *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge: Cambridge University Press.

Hansch, C. (1969). Quantitative approach to biochemical structure-activity relationships. *Accounts of Chemical Research*, *2*(8), 232-239. https://doi.org/10.1021/ar50020a002

Hansch, C. (2003). QSAR and the role of luck in research. *Il Farmaco, 58*, 621-623. https://doi.org/10.1016/s0014-827x(03)00175-7

Hansch, C. (2011). The advent and evolution of QSAR at Pomona College. *Journal of Computer-Aided Molecular Design*, *25*(6), 495-507. https://doi.org/10.1007/s10822-011-9444-y

Hansch, C., & Anderson, S. M. (1967). The structure-activity relationship in barbiturates and its similarity to that in other narcotics. *Journal of Medicinal Chemistry*, *10*(5), 745-753. https://doi.org/10.1021/jm00317a001

Hansch, C., & Fujita, T. (1964). p-σ-π Analysis: A method for the correlation of biological activity and chemical structure. *Journal of the American Chemical Society*, *86*(8), 1616-1626. https://doi.org/10.1021/ja01062a035

Hansch, C., Maloney, P. P., Fujita, T., & Muir, R. M. (1962). Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients. *Nature*, *194*(4824), 178-180. https://doi.org/10.1038/194178b0

Hansch, C., Steward, A. R., Anderson, S. M., & Bentley, D. (1968). The parabolic dependence of drug action upon lipophilic character as revealed by a study of hypnotics. *Journal of Medicinal Chemistry*, *11*(1), 1-11. https://doi.org/10.1021/jm00307a001

Haueis, P., & Kästner, L. (2022). Mechanistic inquiry and scientific pursuit: The case of visual processing. *Studies in History and Philosophy of Science*, *93*, 123-135. https://doi.org/10.1016/j.shpsa.2022.03.007

Hesse, M. B. (1970). *Models and Analogies in Science* (2nd ed.). Notre Dame: University of Notre Dame Press.

Jolles, G., & Wooldridge, K. R. H. (Eds.). (1984). *Drug Design: Fact or Fantasy?*. New York: Academic Press.

Knuuttila, T. (2005). Models, representation, and mediation. *Philosophy of Science*, *72*(5), 1260-1271.

https://doi.org/10.1086/508124

Knuuttila, T. (2011). Modelling and representing: An artefactual approach to model-based representation. *Studies in History and Philosophy of Science Part A*, *42*(2), 262-271. https://doi.org/10.1016/j.shpsa.2010.11.034

Knuuttila, T., & Boon, M. (2011). How do models give us knowledge? The case of Carnot's ideal heat engine. *European Journal for Philosophy of Science*, *1*(3), 309-334. https://doi.org/10.1007/s13194-011-0029-3

Laudan, L. (1977). *Progress and its problems: Towards a theory of scientific growth*. Berkeley: University of California Press.

Lesch, J. E. (2008). 2008 Kremers Award Lecture: Dreams of reason: historical perspective on rational drug design. *Pharmacy in History*, *50*(4), 131-139.

MacLeod, M. (2016). Heuristic approaches to models and modeling in systems biology. *Biology and Philosophy*, *31*(3), 353-372. https://doi.org/10.1007/s10539-015-9491-1

Martin, Y. C. (1978). *Quantitative Drug Design: A Critical Introduction*. New York: Marcel Dekker.

Martin, Y. C. (2012). Development of QSAR. In D. J. Livingstone, & A. M. Davis (Eds.), *Drug design Strategies: Quantitative Approaches* (pp. 60-87). Cambridge: The Royal Society of Chemistry.

Martin, Y. C. (2018). How medicinal chemists learned about log P. *Journal of Computer-Aided Molecular Design*, *32*, 809-819. https://doi.org/10.1007/s10822-018-0127-9

McMullin, E. (1976). The fertility of theory and the unit for appraisal in science. In R. S. Cohen, P. K. Feyerabend, & M. W. Wartofsky (Eds.), *Essays in Memory of Imre Lakatos. Boston Studies in the Philosophy of Science* (Vol. 39, pp. 395–432). Dordrecht: Reidel.

Morgan, M. S., & Morrison, M. (Eds.) (1999). *Models as Mediators: Perspectives on Natural and Social Science.* Cambridge: Cambridge University Press.

Nickles, T. (2006). Heuristic appraisal: Context of discovery or justification?. In J. Schickore, & F. Steinle (Eds.), *Revisiting Discovery and Justification: Historical and Philosophical Perspectives on the Context Distinction* (pp. 159-182). Dordrecht: Springer.

Nyrup, R. (2015). How explanatory reasoning justifies pursuit: A Peircean view of IBE. *Philosophy of Science*, *82*(5), 749-760. https://doi.org/10.1086/683262

Nyrup, R. (2020). Of water drops and atomic nuclei: Analogies and pursuit worthiness in science. *The British*

Journal for the Philosophy of Science, *71*(3), 881-903. https://doi.org/10.1093/bjps/axy036

OECD (2021), *OECD Science, Technology and Innovation Outlook 2021: Times of Crisis and Opportunity*. Paris: OECD Publishing. https://doi.org/10.1787/75f79015-en

Popper, K. (1959 [1934]). *The Logic of Scientific Discovery*. London: Hutchinson.

Purcell, W. P., Singer, J. A., Sundaram, K., & Parks, G. L. (1970). Quantitative structure-activity relationships and molecular orbitals in medicinal chemistry. In A. Burger (Ed.), *Medicinal Chemistry* (3rd ed.) (pp. 164-192). New York: Wiley-Interscience.

Quirke, V. (2006). Putting theory into practice: James Black, receptor theory and the development of the beta-blockers at ICI, 1958–1978. *Medical history*, *50*(1), 69-92. https://doi.org/10.1017/S0025727300009455

Schummer, J. (2021). Knowing-through-Making in Chemistry and Biology: A Study of Comparative Epistemology. *HYLE–International Journal for Philosophy of Chemistry*, *27*, 117-142.

Selassie, C., & Verma, R. P. (2010). History of Quantitative Structure-Activity Relationships. In D. J. Abraham, & D. P. Rotella (Eds.), *Burger's Medicinal Chemistry, Drug Discovery, and Development* (7th ed.) (pp. 1-96). John Wiley & Sons, Inc.

Šešelja, D., & Straßer, C. (2014). Epistemic justification in the context of pursuit: A coherentist approach. *Synthese*, *191*(13), 3111-3141. https://doi.org/10.1007/s11229-014-0476-4

Šešelja, D., Kosolosky, L., & Straßer, C. (2012). The rationality of scientific reasoning in the context of pursuit: Drawing appropriate distinctions. *Philosophica*, *86*, 51-82.

Shan, Y. (2020). *Doing Integrated History and Philosophy of Science: A Case Study of the Origin of Genetics*. Cham: Springer. https://doi.org/10.1007/978-3-030-50617-9

Simon, J. (2012). Chemistry and Pharmacy: A Philosophical Inquiry into an Evolving Relationship. In A. I. Woody, R. F. Henry, & P. Needham (Eds.), *Handbook of the Philosophy of Science* (Vol. 6) (pp. 519-530). North Holland: Amsterdam.

Unger, S. H. (1980). Consequences of the Hansch paradigm for the pharmaceutical industry. In E. J. Ariëns (Ed.), *Medicinal Chemistry* (Vol. 9) (pp. 47-119). New York: Academic Press.

Veldstra, H. (1953). The relation of chemical structure to biological activity in growth substances. *Annual Review of Plant Physiology*, *4*(1), 151-198.

Woodward, J. F. (2011). Data and phenomena: A restatement and defense. *Synthese*, *182*(1), 165-179.

https://doi.org/10.1007/s11229-009-9618-5