DISS ETH NO. 29595

Embodied Quantum Chemistry Learning from Haptic Feedback

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES
(Dr. sc. ETH Zurich)

presented by
Charlotte H. Müller

MSc Computational Science, University of Zurich

born on 19.04.1995

accepted on the recommendation of

Prof. Dr. Markus Reiher, examiner
Prof. Dr. Manu Kapur, co-examiner

2023

Charlotte H. Müller:
Embodied Quantum Chemistry Learning from Haptic Feedback
Dissertation ETH Zurich No. 29595, 2023.

[...], the things you can see, feel, smell,
that's what's real, what else do you
need to know, which means that the
things I could see with my own eyes
had to be the truth, yet another truth
to add to all the other truths I'd been
oblivious to before, [...]
from *A Girl Called Eel* by Ali Zamir

Contents

Abstract

Chemistry in general and quantum chemistry in particular are difficult to learn due to many domain-specific concepts such as chemical bonds or electronic energy being imperceptible. Therefore, learners cannot directly connect these concepts to prior experience. It was demonstrated that haptic feedback has the potential to be an effective facilitator for learning of such scientific concepts. However, while overall, emotional results are positive, cognitive or performance-related results remain ambiguous. Furthermore, little is known about the underlying salient learning mechanisms which are activated when learning from haptic feedback. That is to say, it is unclear *how* haptic feedback facilitates learning.

In this thesis, we present a learning environment that consists of a graphical as well as a haptic user interface and that exploits multiple multimodal representations for embodied learning. We present four studies. First, bachelor students tested the environment in a usability study. The optimized learning environment was then applied in two learning studies, in which we investigated the effect of haptic feedback on salient learning mechanisms and learning outcome in second- and fourth-semester students. Finally, the results of these studies inspired a follow-up study, in the scope of which we investigated the effect of explicit metaphorical explanation on the outcome. The learning studies all followed a problem-solving prior to instruction design. In contrast to instruction-first approaches, the problem-solving allowed the students to first interact with and reflect on the target concept before receiving explicit instruction.

We find that haptic feedback is hindering if the embodied experience is not successfully mapped to the target concept by the students. We attribute this effect to the distraction from the available and generally well understood visual representations by the haptic feedback. We argue that by receiving an explicit explanation of the target conceptual metaphor *Reactions are Hiking over a Mountain*, the students will benefit more from the haptic feedback. This hypothesis is reinforced by increased curiosity, positive affect and perceived relevance observed within the group that participated in an embodied experience coupled with such a metaphorical explanation. The difference in performance was not statistically significant.

Finally, the aggregated qualitative data of all studies provide a summary of student conceptions of the chemical bond concept. We find that students mostly conceptualized

the chemical bond as physical entity or related to energy. We further conclude that the students encounter difficulties in understanding the contextual dependence of the decision when to utilize which model.

Zusammenfassung

Chemie im Allgemeinen und Quantenchemie im Besonderen sind schwer zu erlernen, da viele chemiespezifischen Konzepte wie chemische Bindungen oder elektronische Energie nicht wahrnehmbar sind. Daher können die Lernenden diese Konzepte nicht direkt mit früheren Erfahrungen in Verbindung bringen. Es wurde gezeigt, dass haptisches Feedback das Potenzial hat, das Erlernen solcher wissenschaftlicher Konzepte wirksam zu unterstützten. Während jedoch die emotionalen Ergebnisse insgesamt positiv ausfallen, bleiben die kognitiven oder leistungsbezogenen Ergebnisse ambivalent. Darüber hinaus ist wenig über die zugrunde liegenden Lernmechanismen bekannt, welche beim Lernen mit haptischem Feedback aktiviert werden. Das heisst, es ist unklar, *wie* haptisches Feedback das Lernen erleichtert.

In dieser Dissertation stellen wir eine Lernumgebung vor, welche sowohl aus einer grafischen als auch aus einer haptischen Benutzeroberfläche besteht und die mehrere multimodale Repräsentationen für das verkörperte Lernen nutzt. Wir stellen vier Studien vor. Zunächst testeten Bachelor-Studierende die Umgebung in einer Usability-Studie. Die optimierte Lernumgebung wurde dann in zwei Lernstudien angewandt, in welchen die Auswirkung des haptischen Feedbacks auf die wichtigsten Lernmechanismen und den Lernerfolg von Studierenden des zweiten und vierten Semesters untersucht wurden. Schliesslich führten die Ergebnisse dieser Studien zu einer Folgestudie, in welcher die Wirkung expliziter metaphorischer Erklärung auf den Lernerfolg untersucht wurde. Die Lernstudien folgten alle einem *Problemlösen-vor-Instruktion* (PS-I) Design. Im Gegensatz zu Ansätzen, bei welchen die Unterweisung im Vordergrund steht, konnten die Studierenden beim Problemlösen zunächst mit dem Zielkonzept interagieren und darüber reflektieren, bevor sie explizit unterrichtet wurden.

Wir stellen fest, dass haptisches Feedback hinderlich ist, wenn das verkörperte Erlebnis von den Studierenden nicht erfolgreich auf das Zielkonzept abgebildet wird. Wir führen diesen Effekt auf die Ablenkung von den verfügbaren und allgemein gut verstandenen visuellen Repräsentationen durch das haptische Feedback zurück. Wir argumentieren, dass Studierende mehr vom haptischen Feedback profitieren würden, wenn sie die konzeptionelle Zielmetapher *Reaktionen sind Wanderungen über einen Berg* explizit erhalten würden. Diese Hypothese wird durch gesteigerte Neugierde, positiven Affekt und wahrgenommene Relevanz gestützt, welche in der Gruppe beobachtet wurden, die eine

verkörperte Erfahrung zusammen mit einer solchen metaphorischen Erklärung erhalten haben. Der Unterschied bezüglich des Lernerfolgs war statistisch nicht signifikant.

Schliesslich bieten die aggregierten qualitativen Daten aller Studien eine Zusammenfassung der die chemische Bindung betreffenden Konzeptionen von Studierenden. Wir stellen fest, dass die Studierenden die chemische Bindung meist als physikalische Einheit oder im Zusammenhang mit Energie betrachten. Wir kommen ferner zum Schluss, dass Studierende Schwierigkeiten haben, die Kontextabhängigkeit der Entscheidung, wann welches Modell verwendet werden soll, zu vestehen.

1

Introduction*

The word 'haptics' originates from the Greek words *haptikos*, which means 'able to touch' and *haptesthai*, which means 'able to lay hold of'.[1,2] In research, haptics encompasses the study of touch, which can either be through tactile interaction – temperature, texture, weight – or through force feedback. In this work, we are interested in how force feedback can be exploited to ground or map abstract and intangible concepts to bodily experiences.[3,4]

The framework of grounded and embodied learning (GEL)[3] argues that this mapping onto a familiar domain allows the learner to make connections to prior knowledge and experiences and therefore better integrate the new concepts in existing knowledge. In chemistry specifically, learners are often confronted with either imperceptible or fully symbolic concepts that they then somehow have to integrate into their prior knowledge. Not only are these concepts themselves hard to grasp, perceivable concepts such as colour or density emerge from these inherently abstract concepts.[5,6] Colour or density have no meaning on a molecular level; however, the molecular constitution dictates them. We elaborate on the theoretical framework of this work in chapter 2. Specifically, we will position the chemical knowledge structure[5,7] in the framework embodied cognition[8] and discuss implications of this framing to grounded and embodied learning.[3] Furthermore, we present relevant design principles for learning environments includ-

*This chapter is reproduced in part with permission from C. H. Müller, M. Reiher, M. Kapur *J. Comput. Assist. Learn.* **In press**.

ing multiple multimodal representations as well as the design concept for the learning studies, preparation for future learning (PFL).[9]

Previous empirical research on haptic learning environments supports the theoretical argument that abstract concepts can be made accessible to the senses through haptic feedback and that therefore, the latter has the potential to be an effective facilitator for learning.[4,10] While it is impossible to feel the actual force operating on atoms, a meaningful metaphor can be constructed. Haptic learning environments traditionally consist of a simulation on a computer screen which is navigated with a haptic device that allows for haptic feedback in form of repulsion and attraction. In a systematic review for K-12 education, Zacharia[4] found ambiguous results for 12 studies comparing virtual manipulatives with and without the provision of touch sensory (haptic) feedback. He conjectured that haptic feedback might be beneficial for grounding concepts that cannot be captured visually, but stresses the need for more research as the sample sizes of the discussed studies are quite low. He concluded that "we are still missing most of the information needed for developing a framework depicting *when*, *what* and *how* touch sensory feedback should be provided to learners" (p. 135). The literature review of interactive quantum chemistry developments, empirical evidence of learning from haptic feedback as well as of misconceptions that university students hold around relevant concepts such as energy, chemical bonding processes and the specific reaction mechanism that we used as example problem, nucleophilic substitution reactions, is presented in-depth in chapter 3.

Finally, we present the studies conducted in the scope of this project. We optimized the interactive user interface SCINE Heron based on findings from a usability study presented in chapter 4. For this purpose, we introduced undergraduate students to a real-time quantum chemical simulation that updates the positions of the atoms according to the actions of the learner on the position of a manipulated atom.[11–14] Specifically, the learner can choose one atom with a computer mouse or haptic device[15] and thereby manipulate the molecule. By virtue of continuous quantum chemical structure optimization, a more stable molecule would follow the selected atom while an unstable molecule would follow less resulting in the breaking of the bond between the picked atom and the rest of the molecule. Next to technical issues found by the participants, the feedback focused on an improved three-dimensional aid and hiding unnecessary settings for learners.

The effect on the learning process was then explored in two problem solving followed by instruction[9,16–19] (PS-I) studies, in which the interactive simulation was employed in the problem-solving phase (see chapter 5 for the study with fourth-semester students and chapter 6 for the study with second-semester students). In these studies, we intro-

duced a haptic learning environment to undergraduate students in a problem-solving setting targeting potential energy, its dependence on molecular structure and on the associated forces. The students completed three tasks either with a haptic force-feedback device or a computer mouse followed by an instructional video. Based on previous research on PS-I, we hypothesized that exploration of chemical systems in this environment would increase the students' knowledge gap awareness, positive affect and state curiosity and hence would prepare them for learning from a video (chapter 5) or lecture (chapter 6).[19] In the context of higher education, we attempted to build upon previous findings of increased learning outcome of interatomic forces for students who received haptic feedback.[20] Zohar and Levy's study followed a similar design but targeted high school students' learning of distance dependence of force between two atoms rather than the dependence of force on the electronic energy change and hence molecular structure change. In their learning environment, the students saw two atoms, an energy plot showing the Lennard-Jones potential and the force corresponding to the repulsion present when an atom was moved too close to the other one. Zohar and Levy found a positive effect of their environment on high-school students. In SCINE Heron, the university students further saw an energy graph, now corresponding to the current total electronic energy of a molecular system calculated with highly parameterized quantum chemical methods.[13,21] This results in a higher level of physical realism that becomes relevant for students who have the required mathematical knowledge. Moreover, we extended Zohar and Levy's results by exploring different learning mechanisms quantitatively through questionnaires as well as by assessing the students' learning trajectories qualitatively through open-ended questions. With this mixed-methods design, we explored the extent to which haptic feedback can facilitate learning in our specific setting and for whom this might be the case.

For the fourth-semester students, we found that the haptic feedback had a small negative effect on the learning outcome. We associated this effect on the implicitness of the target conceptual metaphor which might have prevented the students from successfully mapping their prior bodily experiences unto the quantum chemical domain (see chapter 5). Based on this conjecture, we designed a follow-up study that added an explicit explanation of the metaphor to the instructions of the study. Specifically, we compared students who received haptic feedback together with this explanation to students who received the traditional visual representation of forces as arrows together with a literal explanation of what the arrow length corresponds to. We found that perceived relevance, curiosity as well as positive affect were increased in students who received haptic feedback. Chapter 7 describes this follow-up study in more detail.

One aspect of initial quantum chemistry learning is the process of moving from classical

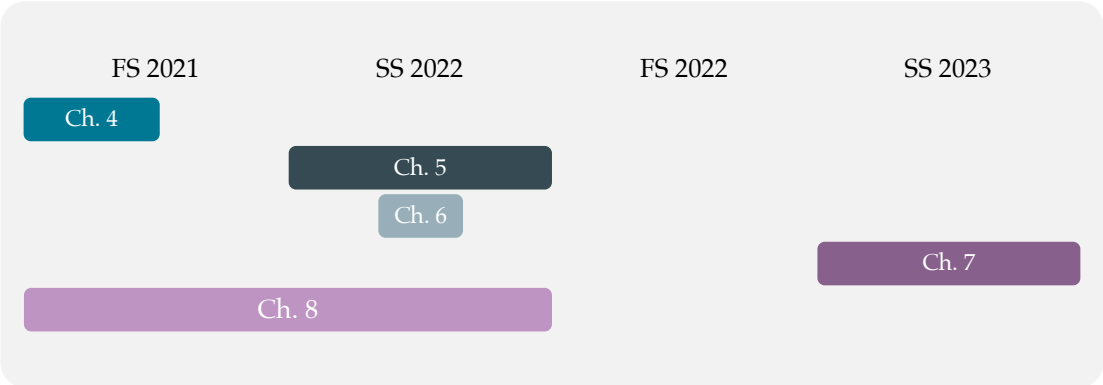


Figure 1.1: Chronological overview of studies conducted in the scope of this dissertation. Chapter 4 presents a usability study while chapters 5, 6 and 7 present PFL studies. Chapter 8 triangulates the qualitative data of the usability and learning studies to present a summary of chemical bond conceptions.

conceptions to quantum conceptions. Many chemistry-specific concepts are introduced highly simplified in a classical context and these simplifications are only later explained or resolved. Student conceptions that lie at either end of this process have been researched well. For instance, in the context of discussing challenges related to chemical bonding, research centers on the different kinds of bonding, ionic, covalent or metallic.[22,23] From this perspective, the chemical bond is mostly conceptualized as a construct of sharing particles between or across nuclei. They are mostly represented with a line or two dots to indicate the location of these quasi-electrons.[24] In contrast, quantum chemistry education research focuses strongly on the conceptions around atomic and molecular orbitals or valence bond theory and hybridization.[25–27] Multiple misconceptions have been identified, mainly around the connection between the physical electron and the mathematical wavefunction, the derivation of molecular from atomic orbitals and the role of hybridization.[24,28] However, little research has been conducted on the influence that these prior conceptions from introductory chemistry have on the learning process in quantum chemistry.

In chapter 8, we explore how the different models were (simultaneously) activated in the specific context of initial quantum chemistry learning. We discuss the influence of learning about quantum chemistry on the understanding of the chemical bond. To this end, we triangulate data from studies with first (chapter 4), second (chapter 6) and fourth (chapter 5) semester students. While the primary target of these studies was not to elicit different chemical bond conceptions, the students still produced data on them in all studies. By analyzing their utterances and written statements on the chemical bond, we could describe their reasoning around it in the specific context of software-assisted initial quantum chemistry learning.

The thesis concludes with a summary of the most important findings and implications of this work. Figure 1.1 shows the content of this dissertation in a chronological manner. In summary, the usability study presented in chapter 4 informed the design and improved the learning environment applied in chapters 5 and 6. The results presented in chapter 5 informed the design of chapter 7. Finally, all qualitative data were re-analyzed for an in-depth account of chemical bond conceptions presented in chapter 8.

The study designs and analysis plans of the studies described in chapter 5 [†], 6 [‡] and 7 [§] were preregistered on the Open Science Framework (OSF) and the data as well as the analysis scripts that support the findings are also openly available in the linked OSF directories. All studies were approved by the ethics committee of ETH Zurich as proposal EK-2021-N-41. Finally, the studies conducted in the scope of this thesis were all done in person. However, they all took place between the years 2021 and 2023 and were therefore situated during the Covid19 pandemic. While we were lucky to be able to conduct studies in person, different measures were active for different studies (gloves and masks for chapter 4, no touching of the keyboard by the researcher during the intervention for chapters 5 and 6 and no explicit measures for chapter 7). Additionally, this pandemic put a strain on the students as well as on the researchers and we must assume that this potentially affected stress levels and anxiety.[29]

[†] https://osf.io/t9wne/?view_only=3c6fdb25cf6947688f56698239f6b2d4, accessed 2023-08-04
[‡] https://osf.io/97yrd/?view_only=4f14de8516b84def9da5aaf04fac8014, accessed 2023-08-04
[§] https://osf.io/c23rx/?view_only=54c73f4b83fa48ac87b35f40f9ff879d, accessed 2023-08-04

2

Embodied Chemistry Learning*

Throughout this work, we argue in the framework of embodied cognition[8] and specifically embodied learning.[3] In the following, we aim to position chemical knowledge in that framework and distinguish between conceptual knowledge and external representations of it. The latter is particularly important as many concepts in chemistry are concerned with sub-microscopic phenomena that are not experiential and hence need to be represented figuratively. In section 2.2, we turn to the educational context that this research was conducted in and further elaborate on the implications of the embodied understanding of cognition on learning in general and chemical understanding specifically.

2.1 Embodied understanding of chemical concepts

We depict the theoretical framework in Figure 2.1. In the following, we begin by introducing embodied cognition,[8] the underlying theory – or theories as we will see – of cognition that we are applying throughout this work. This theory allows us to investigate how humans, specifically students, make sense of abstract concepts through conceptual metaphor, that is how a mapping between the experiential world of representations, symbols and instruments and the imperceptible, sub-microscopic domain of molecules and atoms is necessary in order to understand the latter. We embed the multilevel framework of chemistry suggested by Johnstone[5] into that underlying theory

*This chapter is reproduced in part with permission from C. H. Müller, M. Reiher, M. Kapur *J. Comput. Assist. Learn.* **In press**.

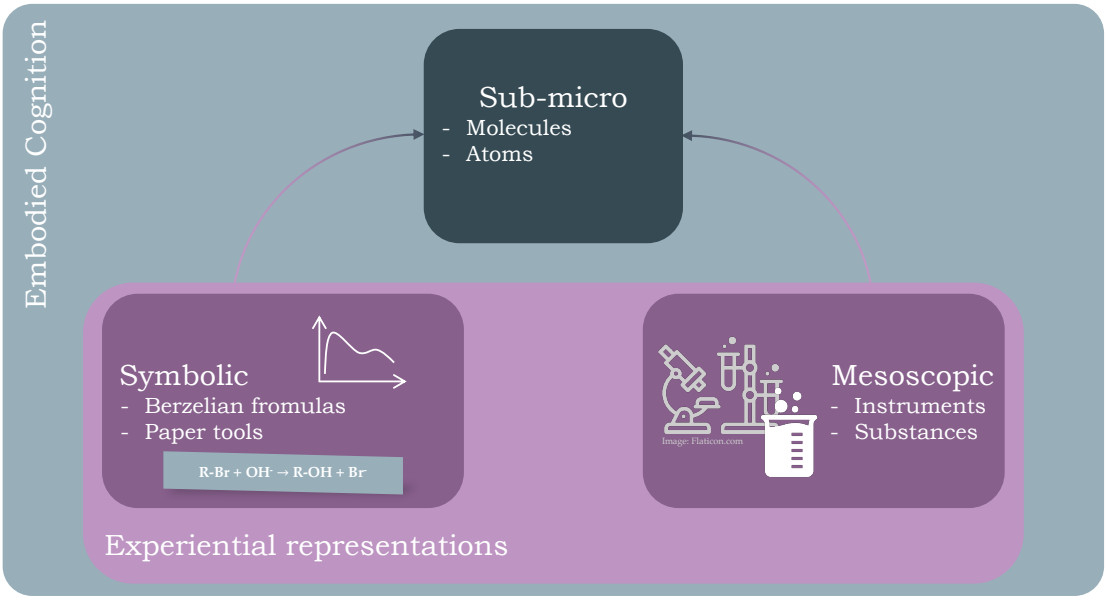


Figure 2.1: The theoretical framework applied in this work, strongly inspired by the multilevel framework of chemistry by Johnstone[5] as well as Klein's interpretation of the role of instruments and pen-and-paper models in chemistry.[7] We argue in the framework of embodied cognition[8] and hence assume that all concepts and representations must be understood in an embodied matter. Metaphorical mappings between experiential domains and the abstract domain are indicated with arrows.

and further analyze what the roles of the three levels symbolic, sub-microscopic and mesoscopic are. We adopt the term mesocosm as introduced by Vollmer[30] to describe what Johnstone called the macroscopic level.[5] In that context, the mesocosm includes the magnitudes perceivable to humans, *i.e.* seconds to decades, millimeter to kilometer, 0°C to 100°C. This section closes with a remark on the ambiguity of *concreteness* in that context as research in the past aimed to organize chemical knowledge into a spectrum from abstract to concrete.[31] We claim that apart from the hierarchy suggested by Lin and co-workers,[31] there are at least two more interpretations of concreteness to consider when designing interventions in chemistry education and that this should be taken into account when comparing interventions based on concreteness.

2.1.1 Embodied cognition

Embodied cognition is a research program that challenges basic assumptions from standard cognitive science. Traditionally, computational cognition as defined in the information processing theory[32] assumes that cognition is an algorithmic process on amodal, abstracted symbols in the brain which is influenced by but strictly separated from the environment. Therefore, cognition itself is removed from perception and action.[33] Along this argumentation, Newell and Simon tried to create a computer program, the general

problem solver (GPS), which solves problems as a human would do.[34] GPS is successful for problems requiring a means-end analysis, such as chess or solving an equation; however, it is also limited to them. The assumption of algorithmic cognition is challenged by the observation that often, action is an important part of creating inputs to cognition. Examples include raising from a chair to see something blocked by an obstacle or picking up an item to inspect it from different perspectives. If the creation of inputs requires active exploration of the environment, the gears of cognition must not only lie in the brain, but also in other organs and even outside of bodies.[8,35] This realization poses further questions connected to the role of (human) bodies in cognition, questions that the research program of embodied cognition aims to answer.[8]

Shapiro suggested to call 'embodied cognition' a research program as there are multiple schools of thought that reject the notion of computational cognition to different extents[36] such as the embodied mind,[37] dynamical systems theory[38] or extended cognition.[35] Shapiro[8] attempted to organize the different theories into what he calls hypotheses, which we present in the following and which form the foundation of our understanding of embodied cognition.

The three hypotheses can be summarized as follows.

- The **conceptualization** hypothesis suggests that "to conceive of the world as a human being does requires having a body like a human being's" (p. 71 of Ref. 8). Throughout this work, we employ the plural *bodies* to acknowledge the diversity of human bodies.[39]
- The **replacement** hypothesis suggests that "[...] interaction with [the organism's] environment replaces the need for representational processes thought to have been at the core of cognition" (p. 4 of Ref. 8) such as mental algorithms.
- The **constitution** hypothesis suggests that "[...] the body is, literally, part of the mind. More radically, [...] the mind extends even beyond the body, into the world" (p. 159 of Ref. 8).

To go into depth of all arguments made by proponents of these hypotheses is beyond the scope of this work. However, we elaborate on two points fundamentally important for embodied learning as described in section 2.2.

First, the proponents of the conceptualization hypothesis, Lakoff and Johnson, argued that how bodies of humans are shaped and function form how we can conceptualize and categorize the world.[40] Specifically, they argued that by constructing mappings between a known and familiar domain and a more unfamiliar domain, we can create meaning and understanding of this unfamiliar domain. For example, the concept of *relationship* might be hard to understand at first, but in terms of the metaphor *relationships as journey*, all kinds of concepts from the domain *journey* can be mapped onto the domain *relationship*,

such as travelling together, standing at a crossroad, sailing smooth or going separate ways.[8] We expand on what this means for chemistry learning in subsection 2.1.3.

Lakoff and Johnson's conceptual metaphor theory has further implications on how we conceptualize the chemical field. Scientific knowledge is inherently created through human actions[41] and one main question of the philosophy of chemistry is what and how "linguistic (or theoretical) actions and actions on the laboratory scale" lead to chemical knowledge.[42] This quote was translated from German and what is interesting about this wording is the categorization of language as an action. Not only are the results created in the laboratory of importance to shape the chemical field, but equally as important is the terminology that we choose to describe this research with. If we argue in the framework of embodied cognition, meaning is created through describing chemical processes in terms of models that map experiential representations onto the concept of interest (see Figure 2.1). For example, one of the most fundamental concepts in chemistry, the molecular structure, was defined metaphorically through the spring-and-ball model by Kekulé.[43,44] If we argue in favor of the conceptualization hypothesis, these metaphors are dependent on the human bodies. It would be interesting to pursue how being aware of this and potentially altering our bodies (such as in immersive virtual reality) might influence chemical reasoning. A similar proposition was made in mathematics with semantic avatars.[45]

Second, gestures are an excellent example of how bodies can support cognition and illustrate the implications of the departure from a purely computational model for thinking. Shapiro argues that as we gesture not only in communication when the other party is present, but also on phone calls or when we are alone, gestures must be more than scaffolds for communication but rather constituents of thinking.[8] Consequently, they are one of the main arguments brought forward by proponents of the constitution hypothesis. This is supported by experimental evidence such as people less fluently describing spatial situations when being prohibited from gesturing.[46] This is relevant to this work for two reasons. First, chemistry is a spatially challenging field and increased spatial ability has been shown to have a positive effect on chemistry performance.[47] Simultaneously, young girls tend to gesture less than boys resulting in them performing worse in mental rotation tasks.[48] While this is not directly under investigation in this work, potential gender differences in spatial ability and hence the spatial aspects of chemistry content was taken into account in the design process of our learning environment in SCINE Heron[49] by attempting to reduce the amount of mental rotation needed to be able to interpret the environment (see chapter 4). Moreover, we measured spatial ability in the interventions described in chapters 4 to 6 in order to assess the interaction effect of condition and spatial ability on the learning outcome. Second, the learning en-

vironment included directed action as opposed to spontaneous gestures. While guided action can have a positive effect on learning processes,[50,51] the loss of agency may also have negative effects by limiting the opportunities for spontaneous gesturing.

2.1.2 The three levels of chemistry knowledge

It was argued that the reason, why science and chemistry specifically are supposedly difficult to learn lies in the inherent structure of chemical knowledge.[5,6,52] Johnstone introduced the widely applied multi-level framework of chemical knowledge that argues that chemists have to understand their craft on three levels – the mesoscopic level, the sub-microscopic level and the symbolic level[5] (see Figure 2.1). To illustrate, picture a chemist working in the laboratory. She sees how the substance in her flask changes – maybe the color changes, maybe a solid forms (mesoscopic level). As a trained chemist, she has her hypotheses about what happens on a molecular scale, which reactants are involved and how they interact with each other (sub-microscopic level). She has written down the reaction mechanism on the white board and has calculated the expected reaction energy and activation barrier (symbolic level). While to her, the connection between these levels is straight forward and she can easily jump between them, a novice has to make these connections actively. The connection between the mesoscopic and sub-microscopic level is not trivial, as chemistry is an emergent science.[44,52] This means that while the properties on the mesoscopic scale (color, state of aggregation) depend on the sub-microscopic scale (the molecules), they do not have meaning on the sub-microscopic scale. A molecule does not have a color or a state of aggregation. This has been argued to be one of the sources for many misunderstandings in students.[6] For example, students might attribute an emergent property of a system to particular components[53–57] or they might view an emergent property of a system as the sum of properties of constituent entities.[56–59]

Niebert and Gropengiesser generalized the argumentation of Johnstone to science education.[60] They argued that as understanding is firmly grounded in experience which has to be on the mesoscopic scale, meaningful conceptual metaphors must be constructed that are sourced in the mesocosm and target the abstract concept on the sub-microscopic or macroscopic scale. Here, grounding denotes the process of mapping "novel ideas and symbols to modality-specific [sensorimotor] experiences that are personally meaningful"[3] and the macroscopic scale includes scales too big for humans to perceive meaningfully, such as the atmosphere or the solar system. The role of metaphors for the understanding of chemical ideas is detailed in the next subsection.

Moreover, the symbolic level in itself is highly complex with one representation bearing multiple meanings.[7] For example, the Berzelian notation (*e.g.* H_2O) serves as i)

abbreviation for names of mesoscopic chemical compounds,[61] ii) referent to empirical, stoichiometric relations[62] and iii) referent to atomic weights.[63] In her book,[7] Klein argues beautifully how the language of chemistry developed and what cultural implications it brings to the field. Specifically, she draws a similar picture of the chemical knowledge structure as Johnstone by distinguishing the Berzelian sign system (on the symbolic level) from the laboratory manipulations and instruments (on the mesoscopic level) while simultaneously stressing the high interconnectivity of the two aspects or levels of chemistry. She adds another important thought to our framework that is especially important when considering how to teach chemistry to novices that might not yet be familiar with the domain-specific language and models. She argues that "[a]lthough both kinds of tools [paper tools and physical laboratory tools] embody intellectual assumptions, they are no longer epistemically relevant in themselves but serve to investigate and represent new scientific objects." (p. 3 of Ref. 7). Therefore, the mesoscopic and the symbolic level can be considered representational levels that serve the indispensable role of representing the sub-microscopic level (see Figure 2.1).

Similarly, Del Re wrote that since humans have "no direct perception of a large majority of the objects which science has detected in the physical universe, [their] knowledge is based on analogies with ad hoc objects".[44] In anticipation of the next subsection, this can be rephrased to the need for metaphors that map features of both tools onto the object under study – molecules and atoms. To further understand how we move between the levels described here from an embodied cognition perspective, we need to introduce the concept of conceptual metaphor which allows us to argue in terms of mapping between imperceptible and perceivable domains.[60]

2.1.3 Chemical understanding via conceptual metaphor

Metaphors play a crucial role in developing, understanding and communicating chemical ideas.[64] Mahootian argued that metaphors provide "a cognitive wedge that opens space for speculation".[64] They allow to express one thing as another and hence map specific properties from one domain onto another. If understood well, this allows for idealized experiments in the sense that an outcome is successfully predicted based on a metaphorical model system.[44] While this language has been proven to be a powerful tool for chemists, students often fall into the trap of interpreting these models, these idealized experiments, as literal representations instead of metaphorical substitutes. Therefore, they are later confused by different, seemingly contradicting models of the same concept.[65] Here, we are focusing on the educational perspective of this aspect but we acknowledge that there exists a rich literature on the historical and cultural role of metaphors in chemistry research and practice.[44,64–66] While research concerned

with teaching and learning focuses on how the correct mapping is constructed based on a defined target concept, these mappings also contribute to the meanings that we assign to it by enabling specific ways of usage, description and understanding of a concept.[8] The language that chemists employ and the metaphors they historically constructed were developed by human minds in a social and cultural context and are consequently limited by these factors. Again, this has implications for research that are, however, beyond the scope for this work. Critically, as the aim of teachers and lecturers is to pass this language on, it is of interest to be aware of its peculiarities and challenges to novices. Moreover, this language was created to a large part by a specific group of people with a certain kind of body. The consequences this has on the learning process of chemical content for example of disabled people remains to be investigated. While Lakoff and Johnson would argue that the impact will be significant,[67] Shapiro disagrees and argues that the same meaning can be generated through different means and hence, one does not necessarily need an identical body to understand meaning assigned to a certain concept by this body.[8] In any case, it is crucial to ensure that the taught conceptual metaphor is mapped correctly by the students. Niebert and Gropengiesser have argued that in science teaching specifically, a mismapping is one of the major sources for misconceptions.[68] Specifically, we focus on the constructionist understanding that misconceptions develop due to metaphorical mismappings between a familiar – or potentially falsely assumed to be familiar by the teacher or lecturer – domains and new chemical ideas.[68]

For this purpose, we apply the definition by Lakoff and Núñez of conceptual metaphor as a cognitive mechanism, a mapping from a familiar (that might imply physical) source domain onto a potentially imperceptible target domain.[69] Knowledge can therefore be interpreted as layers of conceptual metaphors at the core of which the basic concepts[8] or "direct physical experience[s]" (p. 57 of Ref. 67) are located. This is aligned with the conceptualization hypothesis of embodied cognition (see subsection 2.1.1); only what we can experience or what can be mapped metaphorically from what can be experienced can be understood. Some of these basic physical experiences include pushing and pulling. We exploit these experiences to potentially help the students construct conceptual metaphors that map their experiences of pushing and pulling to the unfamiliar domain of quantum chemical forces and potential energy hypersurfaces.

In the following, we briefly diverge to clarify how we conceptualize misconceptions. We are applying the definition by Gilbert and Watts,[70] who state that "'conceptions' be used to focus on the personalized theorising and hypothesising of individuals". They can be accessed through actions of people, including speech, writing, but also non-verbal actions such as gestures. Naturally, not all conceptions are correct or accurate in the sense

that they are not in agreement with current scientific thought (in the case of scientific conceptions). These incomplete or inaccurate conceptions have been termed alternative conceptions[71] or misconceptions.[72] Often, they are accurate in the specific context that they were integrated in but are then wrongfully generalized to other contexts. How learners' conceptions evolve towards expert knowledge, that is how conceptual change can take place is theorized in two major lines of research – theory change[73] and knowledge-in-pieces.[74] Both lines assume that small knowledge structures are crucial for conceptual change. However, theory change treats conceptual change as gradual change of beliefs around a given concept with coherence being high for both expert and novice theories while knowledge-in-pieces assumes that coherence increases with proficiency.[75] Important for this work is the cross-domain mapping that plays a role in both lines of reasoning. This process is characterized by the assumption that new information has to be integrated into prior knowledge. The new information or target domain includes the target concept that is taught. What is included in the source domain depends on the theoretical framework that is applied. However, all mentioned accounts involve minimal knowledge structures that must be – in the framework of embodied cognition – (grounded in) experiential and embodied experiences. Misconceptions arise when these minimal knowledge structures i) are mapped inaccurately to the source domain or ii) are not actually accessible to the learner.[68] For example, Venville and Treagust[76] designed a study that introduced DNA metaphorically as a code, a metaphor that is very commonly applied in textbooks, such as for example in Alberts and co-workers:[77] "cells replicate their DNA [... and] decode the instructions represented in the DNA" (p. 25). They found that whether this metaphor is productive for learning depends strongly on the learner's conception of code. For example, one of their students potentially conceptualized code as a series of numbers and consequently did not conceptualize DNA as a code, but rather as having a code, which results in different, potentially incorrect consequences.[68]

For chemistry specifically, Talanquer[78] has suggested a model to classify and understand students' misconceptions based on what he calls commonsense chemistry. We argue that his argumentation aligns very well with the conceptual change framework described in this section. Students start their academic journey typically with most experience in everyday activities and commonsense reasoning and much less experience in the chemical domain, especially due to the emergent nature of chemical knowledge (see subsection 2.1.2), which can result in misconceptions.[6] Therefore, these commonsense conceptions should be taken into account when introducing new concepts. Students are not presented with new information in a vacuum, but they ground this information in prior experience or knowledge. The lecturer must then provide grounding opportunities to guide this grounding process in a productive direction. Grounding was earlier

described as the mapping of new, potential inexperiential information to meaningful sensorimotor experiences. In that sense, it describes precisely the construction of conceptual metaphors. We discuss the impact of an embodied perspective onto education specifically in section 2.2.

2.1.4 Concreteness in quantum chemistry

Concreteness fading is an instructional approach that gained momentum in mathematics education and has recently found applications in science teaching.[79] Following this approach, instruction initially introduces a concept with concrete representations that are then stripped sequentially of their contextual details until an idealized representation of the concept is obtained. It was found that concreteness fading facilitates learning and transfer.[79] However, the generalizability of this approach was recently questioned by Kokkonen and Schalk[80] due to the dissimilarity of the external representations of the target concept in the different domains and hence the triggered learning processes. In the context of chemistry education, it was suggested that chemical representations could be organized along this spectrum from abstract to concrete.[31] Due to the popularity of the concreteness fading approach together with the somewhat conflicting nature of the presented theoretical framework in this work and the linear framework of Lin and co-workers,[31] we address in the following the verbal dispute[81] that we suspect around what concrete factually means in a chemical context.

In the context of mathematics education, it was suggested that concreteness can be defined along multiple dimensions, specifically sensory accessibility, specificity or familiarity.[82] Here, we are applying this reasoning onto the domain of chemistry to argue why a similar discussion on the multidimensional aspect of concreteness should be had in the chemical domain and what this means for the instructional approach of concreteness fading.

Lin and co-workers[31] defined concreteness as a continuum from mesoscopic, over submicroscopic to symbolic representations. Specifically, they defined the concreteness of a representation as "the similarity to the referent or intended meaning of the representation". With similarity meaning resemblance or correspondence, this definition relies on the underlying assumption that the referent itself is experiential in some form. While this applies in experimental chemistry, where reactions can be observed in the laboratory, this definition is limited in the case of quantum chemistry. For example, it would be hard to concretely represent an electron as a wave-particle since this is inherently intangible and invisible.

Furthermore, we are building on Chatain and co-workers's work by arguing that one

representation can be concrete and abstract at the same time.[82] What matters – in particular in the context of concreteness fading – is the dimension that we are interested in. For example, the potential energy is a multidimensional function that depends on the degrees of freedom of a molecule and is therefore mathematical, symbolic and hence abstract by the definition given above, especially to learners. However, the energy is often visualized as a hypersurface and metaphorically described as a geographical landscape.[83,84] This representation is concrete in the sense that it can be grounded in familiar concepts but it is still general and therefore abstract. Moreover, it does not fulfill the definition by Lin and co-workers[31] as it is not similar to the concept of energy, as energy is not a perceptual concept. For another example, the derivative of the abstract referent energy along one reaction coordinate is proportional to the force. This concept has been introduced as haptic feedback and therefore allowed the learners to construct meaningful metaphors that are highly concrete since they allow grounding in sensory experience.[15] However, again, the derivative of energy is not *similar* to the representation and therefore not concrete in the sense intended by Lin and co-workers.

Importantly for the educational context, different kinds of concreteness can invoke different learning mechanisms.[82] It was found that manipulation concreteness – concretizing through manipulation – evokes the cognitive mechanism of sensorimotor simulation while embodied concreteness – concretizing through grounding in embodied experience – additionally activates the mechanisms direct induction and modal priming.[82,85] The latter improves perceived relevance as found by Chatain and co-workers[82] as well as in this work (see chapter 7) while the former does not necessarily. Therefore, two concreteness fading interventions might activate different learning mechanisms depending on the definition of concrete that is applied.

In conclusion, we suggest that there is a verbal dispute around the concept of concreteness beyond the chemistry triplet that might be an additional obstacle for the comparability of instructional approaches such as concreteness fading.[81] While Lin and co-workers's[31] definition works well for experimental chemistry, we argue that there is a need for expansion of the definition to multiple dimensions if we want to be able to discuss all domain-specific concepts of chemistry, in particular quantum chemistry. In this section, we have presented a framework that allows the separation of the physical sub-microscopic level of chemistry and the levels that serve representational purposes, the mesoscopic and symbolic level. The levels are interconnected through conceptual metaphors that map perceivable concepts to abstract domains. For concreteness fading interventions, it might be beneficial to argue in terms of these metaphors and to move from a more *concrete* source domain – where *concrete* needs to be defined – to a more

abstract source domain. This might be more precise and more customizable to what is concrete to the learner.

2.2 Embodied learning from haptic feedback: Design principles

In the previous section, we presented a framework of cognition in which we situated the chemical knowledge domain. We argued that as the main phenomena of interest in chemistry, molecules and atoms and their properties, are not tangible, visible or otherwise perceivable, conceptual mappings need to be constructed from experiential levels, that is the symbolic level and the mesoscopic level to the concept of interest. In this section, we aim to present the implications that this framework has on how we design for and analyze the learning process of chemical concepts. To this end, we introduce three relevant lines of work that are pivotal for the design of the learning environment (see chapter 4) and the learning studies (see chapters 5 to 7). First, Nathan has developed the grounded and embodied learning (GEL) framework[3] which presents principles and mechanisms crucial for the learning process from an embodied cognition perspective. Second, in chemistry education generally and in our specific learning environment, we are working with multiple and here additionally multimodal external representations. As working memory capacity is highly limited, it is critical to carefully design for these representations and we present the principles that we applied in subsection 2.2.2. Finally, all our learning studies are following the identical instructional design, preparation for future learning (PFL[9]). We are presenting the main design principles and the defense as to why this design was suitable for this work in subsection 2.2.3.

2.2.1 Grounded and embodied learning

The grounded and embodied learning (GEL) framework adopts several of the underlying assumptions of embodied cognition.[3] First, the framework assumes that the mind controls a predictive architecture in the sense that multiple possible outputs are simultaneously activated and updated based on the probability of their correctness. Second, the reality of a person is distorted by sensory experiences, which themselves are a product of the person's environment. Third, memories and hence knowledge is constructed and alterable. One consequence of the last assumption was evident in subsection 2.1.3, in which we mentioned that misconceptions can be conceptualized simply as mismappings between prior knowledge and new minimal knowledge structures that were unsuccessfully attempted to be integrated into the former.

Under these assumptions, GEL is concerned with the limited capacity of working memory and how to support the relevant cognitive processes for learning. Working memory

denotes the part of information that is readily available to us and that can therefore be actively used for cognitive tasks such as learning.[86] Specifically, GEL incorporates the dual-process theory which suggests that there are two main ways of thinking, the intuitive, automatic, implicit, parallel and low capacity, type 1 processing and the deliberate, slower, rule-based, serial and capacity-dependent type 2 processing.[87] As learning is considered type 2 processing, GEL is mainly concerned with optimizing and supporting type 2 processing. GEL introduces principles for designing for an embodied understanding of learning. Learning in this context is defined as "lasting change in behavior".[3]

Table 2.1: Different types of embodiment as defined by Nathan[3] and how cognitive mechanisms associated with learning might be activated for each in the context of our work.

Type of embodiment	Grounding	Offloading
Movement & perception	Metaphor for energy change in terms of perception of force	Move visual information of force to haptic channel
Simulation	Instantiations of metaphor when recalling concept	-

There are many ways how to involve bodies in the learning process. GEL differentiates between four kinds of embodiment, movement and perception, gesture, embodied simulation and materialist epistemology[3] (see Table 2.1 for the relevant connections to this work). In the context of this work, we are targeting embodied simulations and perception. The former describes how human minds have the ability to re-experience sensorimotor processes. Under this assumption, perceiving the force that represents an activation barrier in a chemical reaction should have the potential to be re-experienced in a later situation such as a formal learning setting. Similarly, movement and perception ground the meaning of concepts through embodied, conceptual metaphors.[60,68] Since we understand new concepts through integration into existing prior knowledge or experiences, imperceptible concepts such as energy, molecules or the atmosphere have to be understood metaphorically. Metaphors allow to map familiar embodied experiences such as pushing a door open or handling a container onto abstract concepts such as forces in chemical reactions or understanding climate change[68] (see subsection 2.1.3 for a more detailed account of conceptual metaphors).

Furthermore, Nathan specifies four cognitive mechanisms associated with embodied learning, grounding, offloading, cognitive-sensorimotor transduction and participation.[3] In this work, we focus on grounding and offloading. The former is defined as the process of mapping novel ideas and symbols to modality-specific, personally meaningful experiences and hence allowing the students to ground the abstract concepts of potential

energy in a chemical context in prior bodily experiences of attraction or repulsion (see Table 2.1). The latter refers to ways of how bodies and the environment can reduce the demand on the highly limited cognitive resources. Again, metaphors can deliver means for offloading cognitive operations needed for a new domain – such as chemical reactivity – onto a previously grounded domain – such as overcoming a mountain or pushing a door open.

From this follows that experiencing the concept of force instead of solely studying the equations associated with it will allow the students to ground novel concepts in prior experience. In other words, the students may construct a conceptual metaphor between their sensory experiences and the chemical concepts. This is in contrast with the traditional formalism-first approach to teaching abstract concepts.[88] Rather than introducing the symbols and algorithms associated with a concept first and subsequently establishing understanding, the understanding comes first without presuming the needed domain-specific terminology to describe the concept. In our specific case, the students feel the activation barrier of a reaction as repulsive force and notice that they have to push to overcome this barrier, similarly to having to push a door open in order to enter a room. Only later, they receive the instruction that the force is minus the derivative of the energy, $F = -\nabla E$.

In empirical education research, studies have focused on gestures,[50] larger motor actions[51,89,90] or full-body movement[91] and finally, haptic feedback.[20,92–95] One particularity of the latter is the novelty of haptic devices to learners. While the haptic feedback offers potential grounding opportunities, the extraneous load such an unfamiliar device brings should be considered when designing for haptic feedback. The empirical evidence is discussed in more detail in chapter 3. In the following subsection, concrete design principles for multimodal representations are presented.

2.2.2 Learning from interactive multiple and multimodal representations

Representations in the context of cognitive psychology have two separate meanings. First, representations denote the way in which we store and represent knowledge mentally, the mental representations; second, representations denote the way in which knowledge is represented to others, the external representations. Here, we are interested in how to design external representations. They are defined as "the knowledge and structure in the environment, as physical symbols, objects or dimensions (*e.g.* written symbols, beads of abacuses, dimensions of a graph, *etc.*) and as external rules, constraints or relations embedded in physical configurations (*e.g.*, spatial relations of

written digits, visual and spatial layouts of diagrams, physical constraints in abacuses, *etc.*)."[96]

Davis and co-workers organized haptic external representations specifically in their haptic bridge framework.[97] They divided these haptic representations into four major categories that can be visualized as the four squares in a spectrum from isolated to connected on the one axis and iconic to symbolic on the other. The isolated and iconic representations are called haptic mirrors as they aim to mimic the physical experience. They do not represent conceptual metaphors but introduce the actual concept virtually. For example, a surgical simulator that mimics resistance of skin would be considered a haptic mirror. In contrast, when the representation is symbolic and connects two or more dynamic visualizations, Davis and co-workers call it a haptic bridge. This representation has the potential to scaffold the construction of a conceptual metaphor; however, it is not necessary in order for it to be considered to be a haptic bridge in the sense intended by the framework. The haptic feedback itself can simply amplify the connection between the visualizations without adding any further meaning to the experience. For example, Davis and co-workers presented students with the graph as well as the unit circle representation of the sine function which were connected by a haptic feedback that corresponded to the amplitude of the curve.[97] In that case, the haptic feedback's sole role is to scaffold the connection between the two visualizations. While we aim to scaffold the construction of a meaningful conceptual metaphor, we take into account the role that the haptic feedback plays in the learning environment as a whole, that is in relation to all the given visualizations. Hence, we introduce in the following two lines of research that target not necessarily haptic feedback in learning but multiple[98] or multimodal[99] representations in learning.

In chemistry, learners are confronted with a number of different representations such as energy graphs, molecular structure, symbolic equations and reaction mechanisms.[100] In SCINE Heron,[49] the students are presented with a molecular structure represented with balls and sticks, an energy graph and haptic feedback. Therefore, the environment consists of three representations, two visual and one haptic. Consequently, we need to consider design choices for both multiple and multimodal representations. According to the modality principle of instructional design, learning environments that combine verbal and non-verbal representations are most effective if implemented correctly.[101] This includes following the design principles for (interactive) multimodal learning environments. An overview of the principles together with the corresponding implementations in SCINE Heron are presented in Table 2.2. The design choices are heavily based on the assumption that working memory capacity is highly limited and that new information necessarily has to be integrated and organized in the context of prior experiences

and knowledge. We adopt these principles by *guiding* the students through the activity with detailed instructions, inviting them to *reflect* on the activity through open-ended questions, letting them do the tasks in their own *pace* and introducing *pretraining* with the haptic device before the actual intervention.

Table 2.2: Design principles and corresponding theoretical rationale as defined by Moreno and Mayer[101] with the implementation in SCINE Heron (see chapter 4).

Principle	Rationale	Implementation
Guided activity	Prompts selection, organization, and integration of information	The students received detailed instructions incl. screen shots of the suggested action.
Reflection	Encourages more active organization and integration of information	Instructions included open-ended questions and space for reflection in writing.
Feedback	Reduces extraneous processing by providing correct schemas	No feedback was implemented to be able to understand what the students took away from the environment alone.
Pacing	Allows processing of smaller chunks of information in working memory	The students moved through the tasks in their own pace.
Pretraining	Highlights aspects of prior knowledge to be integrated with new information	The students received pretraining with the haptic device and a simple molecule.

In her review of research on multiple representations, Rau explored the conditions under which multiple representations are beneficial compared to a single representation.[98] In particular, she states that not only is it important to understand how students learn *from* multiple representations, but also how students learn *about* multiple representations, that is how students gain representational competencies. She further concludes that visual[99,102] as well as connectional[102] understanding is necessary in order to learn from multiple representations. Specifically, if the students do not understand and do not get instructional support for the visual representations and the connections among them, multiple representations will not be more effective than a single representation.[98] This may cause what she terms a representation dilemma: "How can students learn new content from visual representations they do not yet fully understand, and — at the same time — learn how visual representations show content they have not yet learned?" (p. 719 of Ref. 98). As a design choice, Rau recommends that each representation needs to provide relevant information that overlaps with other representations but not fully.[98] In our case, we fulfilled this recommendation through the underlying continuous electronic structure calculations. The movement of the atoms, the update of the electronic energy graph and the perceived haptic feedback all showed the same information on force if interpreted correctly. Faster movement, steeper ascend or descend of the energy and stronger haptic feedback all corresponded to a larger force. The representation dilemma

can further be avoided by first allowing intuitive understanding through conceptual metaphor and learning the formal representation after.

The presented literature in this subsection focuses heavily on textual and pictorial representation or textual and auditorial representations. However, we argue that their principles translate to contexts that involve visual and haptic information.

2.2.3 Preparation for future learning

The covered design principles were applied in the development process of the learning environment in SCINE Heron (see chapter 4). This learning environment was applied in learning studies situated in authentic higher education learning settings (see chapters 5 to 7). In these studies, we followed a preparation for future learning (PFL[9]) design. This design paradigm suggests to introduce a preparatory activity prior to direct instruction. The paradigm has been founded under the assumption that often, students do not bring the relevant prior knowledge into the classroom or if they do, that it might not be sufficiently activated for the instruction. The students at the universities of Switzerland have a highly heterogeneous background. Some come directly from high school (75.7% between 2005 and 2020) where they already received specialized education in STEM subjects (19.8%), others might have focused on languages (16.9%) or arts (5.8%). Other students might have decided to obtain a university degree after already completing and working in another profession (3.0%).[103]

Schwartz and Bransford suggested to implement contrasting cases prior to learning from instruction in order to facilitate the generation of differentiated knowledge structures that enable to deeply understand the content of the instruction.[9] The preparatory activity can be designed in many ways, such as reading, worked examples, explanation generation or problem-solving. In this work, we are focusing on problem solving followed by instruction[16,18,104,105] (PS-I). In their review, Loibl and co-workers defined three main cognitive mechanisms that are activated in a PS-I setting, prior knowledge activation, awareness of knowledge gaps and deep feature recognition.[17] They suggest that knowledge acquisition starts with a learner arriving at a problem that requires the target knowledge which is not fully covered by the students' existing prior knowledge. The learner then attempts to solve the problem by activating their prior knowledge that will probably not suffice to solve the problem. Through interaction with the problem, that is through contrasting or solution generation, the students realize the deep features of the problem that they need to consider in their solution, as well as their knowledge gaps. When the learner finally arrives at instruction, they should be aware of what input they need in order to be able to solve the problem. Figure 2.2 depicts how the design was implemented in this work. The instruction differed for the three learning

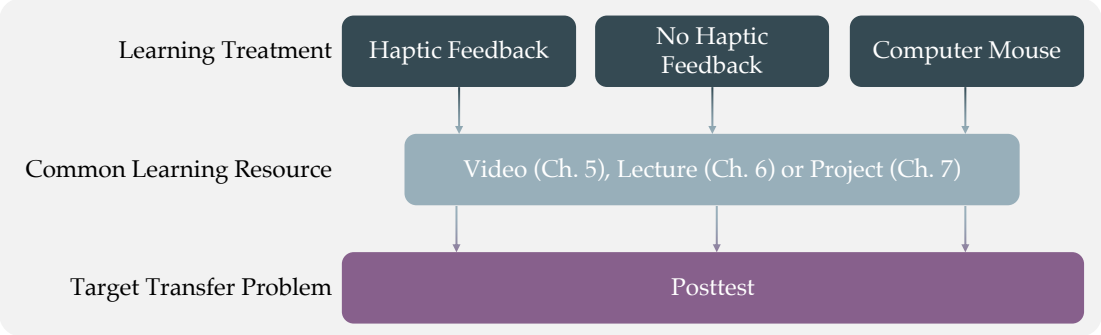


Figure 2.2: The double transfer paradigm as introduced by Schwartz and Martin[104] and as implemented in this work.

studies, but the problem-solving was highly similar with the study of chapter 7 building on the results from chapter 5.

In a meta-analysis, Sinha and Kapur found a significant estimated true effect size (Hedge's g) of 0.87 in favor of PS-I (as compared to I-PS) for conceptual knowledge and transfer after accounting for publication bias.[19] Furthermore, they found instruction building on student solution (Hedge's $g = 0.56, p = 0.02$) and group work as the participation structure (Hedge's $g = 0.49, p = 0.04$) to be highly important predictors of this effect size. Furthermore, empirical results suggest that PS-I potentially positively influences curiosity and affect.[106–110]

Therefore, in the quantitative analysis of the learning studies, we focus on the affective variables curiosity and affect as well as the learning mechanism knowledge gap awareness. Finally, we further measure extraneous load due to the technological novelty of the haptic device.

2.3 Research questions and hypotheses

In this work, we aim to reproduce and build upon previously found positive effects of haptic feedback on learning outcome in the context of initial quantum chemistry learning.[20,111] Therefore, the first research question asks to what extent learning with this specific haptic learning environment facilitates learning of basic quantum chemical concepts compared to the same environment without haptic feedback and the same visual environment navigated with a computer mouse. Moreover, we are interested in how learning mechanisms are influenced by the mode of navigation.

Based on the conceptual framework presented in this chapter, the following hypotheses emerged.

1. We hypothesized posttest performance to be highest for the students receiving hap-

tic feedback (experimental), followed by students who used the computer mouse (control) and last by students who used the haptic device but did not receive haptic feedback (comparison).

2. Based on PS-I literature, we hypothesized positive affect, state curiosity and knowledge gap awareness to be influenced positively by the mode of the PS phase.[19]

These hypotheses are investigated in chapters 5 and 6. In these studies, we found that solely offering grounding opportunities by providing haptic feedback without sufficient support of this grounding leads to a reversed effect of students performing better when they did not receive haptic feedback. We explained this effect with an unsuccessful mapping between the source domain – embodied experiences of repulsion and attraction – and the target domain – forces in a quantum chemical context. This is in alignment with work by Niebert and Gropengiesser, in which they review examples of experiential learning through construction of metaphors and in which they found that unsuccessful interventions could be explained with such a mismapping[60] (see subsection 2.1.3). Hence, we conducted a follow-up study in which we compared students who received haptic feedback together with an explicit explanation of the metaphor to students who saw the forces represented as arrows and an explanation of this representation without any metaphor. Based on previous results and the review by Niebert and Gropengiesser,[60] we hypothesized that

1. *successfully* grounding energy change in prior bodily experience will lead to better performance in a posttest compared to having a visual indication of energy change that is less grounded,
2. bodily grounding leads to an increase in learning mechanisms associated with problem-solving (positive affect, state curiosity and knowledge gap awareness)[19] and
3. due to the meaningful interaction, the haptic feedback group will show higher motivational attributes, specifically attention, relevance, confidence and satisfaction (ARCS[112]).

Hypothesis 1 and 2 are identical to the previous study that lacked support of the metaphorical mapping (see chapter 5). Hypothesis 3 was added based on work by Chatain and co-workers.[82] In a study that targeted the learning of graph theory, they compared abstract, manipulation and embodied interaction and found that the embodied condition scored significantly higher than the others specifically for perceived relevance. They argued that since embodied grounding leads to a stronger connection with the learner's prior experience, the perceived relevance increases.

These hypotheses are investigated in chapter 7.

3

Literature Review of Haptic and Interactive Chemistry (Learning)*

In this chapter, we situate our work in the relevant empirical literature. In section 3.1, we discuss the historical and recent development in interactive quantum chemistry, specifically applications that allow for real-time simulations of molecular systems. In section 3.2, we focus on the educational context and review literature that substantially inspired the software development and learning studies presented in this work. First, an overview of previous haptic chemistry environments is given in section 3.2.1 with targeted attention on educational applications. Second, an account of potential prior conceptions of the students is presented in section 3.2.2. For this work specifically, conceptions of the nucleophilic substitution reaction mechanism, of potential energy and of chemical bonding are important.

3.1 Interactive quantum chemistry

Interactive simulations have a long tradition in chemistry research. In their review, Lanrezac and co-workers define interaction as the ability to modify simulation parameters at run-time or as manipulation of atoms or molecules in the simulated system.[113]

In this work, we first summarize how chemical information is traditionally visualized

*This chapter is reproduced in part with permission from C. H. Müller, M. Reiher, M. Kapur *J. Comput. Assist. Learn.* **In press**.

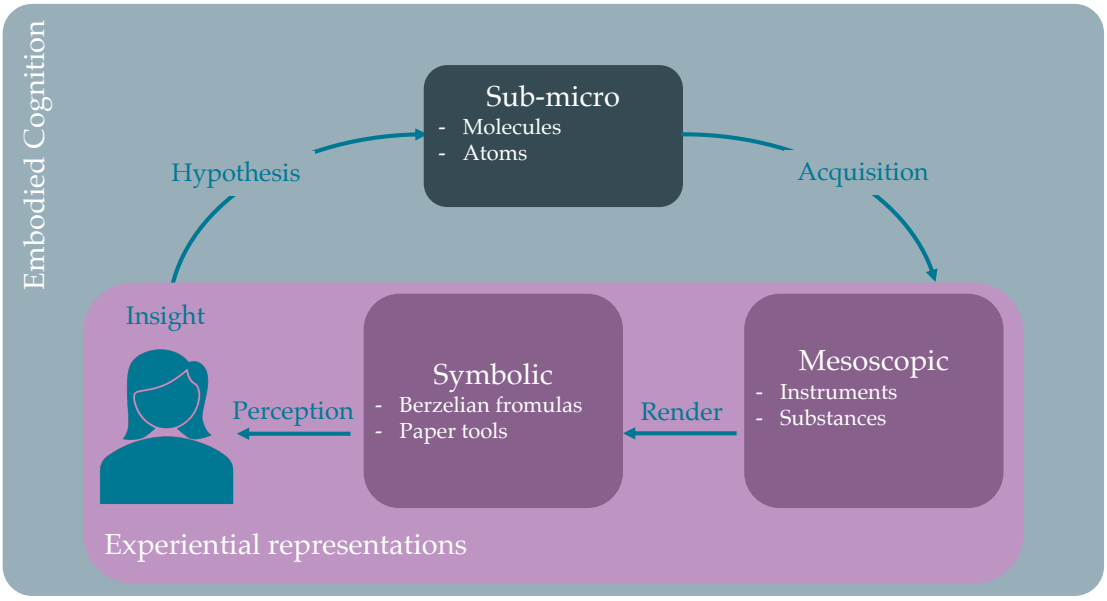


Figure 3.1: The visualization process as suggested by Valle[114] integrated into the conceptual framework from chapter 2. This chapter targets specifically the steps termed 'Acquisition' and 'Render'.

and interacted with and subsequently concentrate more closely on salient interactive chemical applications with emphasis on multimodal representations. For a review of haptic environments specifically for learning, refer to section 3.2.1.

3.1.1 Visualization of chemical information

Valle defines visualization as "the use of computer-supported, interactive, visual representations of data to amplify cognition", where with amplifying cognition, Valle means to ignite imagination and to provide a holistic view.[114] He suggests a framework describing the interaction with (chemical) information, in which the object under study is interacted with in a continuous loop consisting of data acquisition, visualization, perception, insight and new hypothesis forming (see Figure 3.1). Therefore, the form in which we visualize the data is a choice that directly influences the course of this interaction process and hence research itself. Especially calling to mind the crucial role that representation plays in the chemical knowledge structure (see section 2.1), the development of new ways to represent chemical concepts is not only elegant and aesthetically pleasing but more importantly driving scientific thought.

Chemical systems can vary greatly in size and consequently properties that are i) feasible and ii) desirable to visualize. While for proteins, often superficial structures such as potential docking sites are of interest, in reaction mechanisms, the individual atoms

are of importance. Therefore, simplified representations such as ball-and-stick, surface or VDW representations are used depending on the needs of the operator. In the visualizations that allow to distinguish between atoms, the established CPK color scheme is applied.

While this is an obvious statement for chemists, the colors as well as the individual representations, the representational competencies,[98] have to be learned by novices. Once learned, they will prime the viewer for different conceptions. For example, the ball-and-stick representation visualizes the chemical bond as a metaphorical rigid stick that can be broken. This triggers different misconceptions than the VDW representation that does not visualize bonds at all (refer to chapter 8 for a more in-depth analysis of the chemical bond conceptions of ETH students). In his summary comment on the Faraday discussion, Brooks points out that while these visualization techniques are widely applied in all areas of chemistry and chemistry-related fields, a coherent theory for data visualization is lacking.[115] Therefore, the comprehension of how visualization techniques steer hypothesis forming and consequently *doing* and *learning* science is not understood well.

As we discuss further in chapter 5, this assumption that the visualization will not sustainably influence the associations made by the learner, influenced the outcome of the referenced learning study. In this work, we decided to visualize chemical bonds as sticks based on the distance of atoms as this is the most commonly used representation among *chemists*. However, as we discuss later, the sudden appearance of bonds confused *students* as they seemingly contradicted the heuristics held by the students.

3.1.2 Multimodal interactive simulations

An overview of the developments mentioned in this subsection is given in Table 3.1. The first interactive chemical visualization was introduced by Levinthal in 1966 and allowed the operator to manipulate the atoms of a protein.[116] This inspired Brooks to his project GROPE.[117] In addition to the visual interaction, GROPE-III further allows for force feedback. In all described applications, force or haptic feedback refers to the resistance or attraction that a motor inside a haptic device allows for. The haptic device consists of this motor and a connected pen for navigation, which can move in three dimensions. This allows the operator to feel the resistance or attraction in a specific direction in space. Brooks found that chemists were able to reproduce the accurate binding sites in protein-ligand docking mechanisms. These chemists further reported to gain a deeper understanding of the binding site and force field.[117] This confirms a vision presented by Atkinson and co-workers in 1977.[118] They wrote

> Chemists routinely draw potential energy surfaces for molecular internal ro-

> tations, vibrations and collisions. Why not feel these potential surfaces as force fields? Why not feel the dynamics of an atom [...] colliding with another molecule in a chemical reaction? (p. 102 of Ref. 118)

Since GROPE-III, many applications were presented that focused on making different properties accessible to the haptic modality and to further improve algorithms to facilitate that.[119–123] Specifically, the first and still best researched application of haptic molecular simulations lies in molecular docking.[117,124–129] These applications approximate the energy with a sum of electrostatic (Coulombic) and van der Waals interaction energies using Lennard-Jones potentials.

Furthermore, interactive simulation can imply that the molecular system reacts and adapts according to the actions of the operator. This can be achieved by continuously minimizing the energy of the molecular system. The first example in this area is Sculpt, in which small proteins are relaxed according to torsion angles, hydrogen bonds, van der Waals and electrostatic interactions at 11 Hz.[130] In later applications, the frequency of the haptic feedback is increased to allow for smooth perception which requires 1 kHz.[131,132] Applications include force-based approaches[133–136] or semi-empirical approaches for fast visualization of scanning tunneling microscopy images.[137,138]

Another application of interaction in chemistry is the exploration of reactivity and hence the potential energy hypersurfaces. Reiher and co-workers invented the framework of haptic quantum chemistry.[15] While originally, the interaction relied on an interpolation scheme from sets of ab initio data points, the data flow was later improved to allow for an automated and flexible workflow.[139] Further improvements permit on-the-fly single-point calculations,[11,12,140] resulting in the real-time quantum chemistry framework.[11] It was shown that highly parameterized methods such as density functional tight binding methods[141–144] or methods which neglect diatomic differential overlap[145–151] are sufficiently accurate while providing the necessary speed.[140,152] Still, these methods scale with the size of the molecular system. For bigger systems, mediator atomic potentials provide the results between electronic structure results.[13,14,153] This framework is implemented in the graphical user interface SCINE Heron[49] that is further presented in chapter 4. Finally, there have been advances towards training neural networks to learn potential energy surfaces.[154–157]

As our application is a desktop virtual reality environment, it is beyond the scope of this work to review the literature on immersive virtual reality (iVR) and augmented reality (AR) applications here. Reviews on the topic were recently published by Fombona-Pascual and co-workers[158] and Mazzuco and co-workers[159] for chemistry education specifically.

Table 3.1: Overview of the most relevant developments in interactive chemistry applications including method and update frequency. This work implements the algorithmic flow described in Vaucher and co-workers.[13]

Publication	Method	Update rate	Spontaneously updates structure	Remarks
Levinthal, 1966[116]	-	-	Rotation	Structure editor for large molecules
Brooks et al., 1990[117]	Lennard-Jones potential (unspecified)	60 Hz	No	Potentials precalculated on grid points[124]
Surles et al., 1994[130]	User-defined force field	11 Hz (20 residue protein)	Yes	Constrained bond lengths and angles
Levine et al., 1997[125]	Lennard-Jones (unspecified)	-	Yes, not during manipulation	Genetic algorithm
Nagata and co-workers, 2002[126]	GRID potential energy[160–162]	-	No	
Křenek, 2003[119]	Force-field	1 kHz	Yes	Not chemistry specific
Lee, 2004[120]	Lennard-Jones potential (unspecified)	1 kHz	No	Force precalculated on grid points
Wollacott & Merz Jr., 2007[127]	Amber parm94[163]	1 kHz	No	
Marti & Reiher, 2009[15]	Density functional theory (DFT) (BP86[164]/TZVPP[165])	1 kHz	No	Minimal energy path – Interpolating moving least-squares (MEP-IMLS) algorithm
Daunay & Régnier, 2009[121]	Force field (not specified)	1 kHz	Yes, ligand only	
Sourina et al., 2009[128]	OPLS-aa[166] force field	1 kHz	No	Helix-helix docking
Comai & Mazza, 2009[129]	Force field (not specified)	-	No	
Bosson et al., 2013[137]	Atom superposition & electron delocalization molecular orbital (ASED-MO[167]) theory	20 Hz	Yes	Block-adaptive quantum mechanics (BAQM) algorithm
Anthopoulos et al., 2014[133]	MMFF94s force field[168]	-	Yes	GPU tailored algorithm
Vaucher et al., 2016[13]	Semiempirical (PM6[151]) & tight-binding (DFTB[143,144]) methods	1 kHz	Yes	Mediator potential
Iakovou et al., 2017[134]	Gromos54a7 force field[169]	500 Hz	No	GPU tailored algorithm
Jaillet et al., 2017[135]	Interactive modeling universal force field (IM-UFF)	-	Yes	
Dubois et al., 2018[138]	ASED-MO theory + v.d.W. corrections[170]	-	No	Tersoff-Hamann approximation[171]
Matthews et al., 2019[136]	Amber ff14SB force field[172]	500 Hz	Yes	Linear response approach[173]
Amabilino et al., 2019[155]	Interactive ab initio molecular dynamics in virtual reality (iMD-VR[174]) vs. constrained molecular dynamics (CMD)	-	Yes (for iMD-VR)	Neural network approach

3.2 Haptic chemistry learning

For the design of our interactive learning environment, we draw from the empirical literature of haptic chemistry learning environments and the literature on conceptions held by university chemistry students about the target concepts. We elaborate on the relevant findings in the following.

3.2.1 Haptic chemistry learning environments

The empirical evidence of haptic learning environments reviewed by Zacharia shows a positive effect of the haptic feedback for affective and motivational variables.[4] For example, when comparing high school students who received no as compared to full haptic feedback corresponding to the interaction between a virus sample and the probing tip of an atomic force microscope, Jones and co-workers report excitement and increased understanding in all students.[175] However, the evidence for cognitive and hence learning gain is ambiguous. In a review from 2015,[4] seven out of the twelve studies that compare virtual manipulatives with and without haptic feedback find a better understanding of the target concept in students who received haptic feedback (gears,[176,177] cell membrane transport,[178] virus,[179] protein-ligand docking[180,181] and force fields[95]). The studies who report affective results but lack cognitive results mainly argue with the increased cognitive load.[10,182,183] Furthermore, Zacharia conjectures that these mixed results might be attributed either to the failure of physicality to offer fundamental differences to an experience without haptic feedback or to a lack of added value in the haptic experience.[4] For example, in the study by Reiner, the students could only feel a force field. There was no other way to gain access to the information than through the haptic feedback.[95] When asked to draw what they felt, the students drew representations which were highly similar to the ones usually found in textbooks.

More recently, Zohar and Levy found a positive effect of haptic feedback on the learning of the presence of repulsive forces in the chemical binding process.[20] Post hoc comparisons showed that while the three non-haptic conditions (watching a movie, navigating with a computer mouse and navigating with a joy stick) performed comparable, the haptic condition showed significantly bigger learning gain, specifically on the items targeting "separation of atoms increases the potential energy". Zohar and Levy further mention that the haptic feedback makes the force-distance-energy relationships experiential as the haptic feedback makes it difficult to move away from the energetic minimum, highlighting that this minimum represents the most stable form of the two atoms in their example.[20] Similarly, Persson and co-workers report qualitatively that the haptic feedback allowed the students to make connections between energetical and

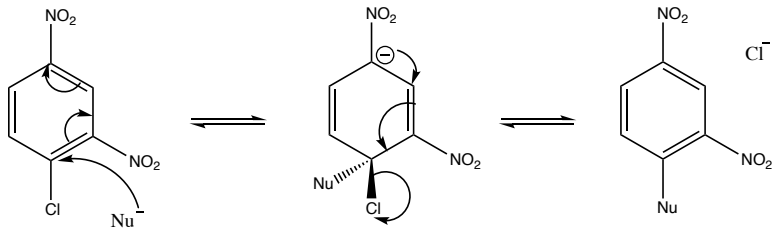


Figure 3.2: Traditionally, reaction mechanisms are presented with the electron-pushing formalism that suggests a goal-oriented movement of the electrons.

dynamic aspects of the simulation.[182] Subsequently, similar settings were explored that combined the haptic learning environment on the chemical bonding process with virtual reality, so far only with affective results.[184,185]

Moreover, the concurrence of the change in molecular structure, force and change in energy was a driving factor. According to students, as reported in the study of Bivall and co-workers, to experience them at the same time was crucial for their understanding.[180] While this point is only a conjecture as they were not able to confirm this quantitatively, it is aligned with the load reducing method *synchronizing* of Mayer and Moreno[186] and illustrates the advantage of interactive simulations over movies and static pictures: The possibility to perceive the force, energy change and structural change simultaneously.

3.2.2 Student conceptions

Several target concepts are of importance in the learning studies described in chapters 5 to 7. In this section, we present empirical research on the misconceptions specifically around the concepts energy, nucleophilic substitution reactions, chemical bonding and orbitals. All presented research was conducted outside of Switzerland and it is possible that there remain cultural and social differences between the research contexts.

The nucleophilic substitution reaction mechanism

Traditionally, chemical structures are represented as Lewis structures with lines for bonds and letters for atoms while reaction mechanisms are indicated with arrows indicating the electron flow, the electron-pushing formalism (see Figure 3.2). This presents the electron flow as deterministic while it neglects activation barriers and therefore repulsive forces.[187] In nucleophilic substitution reactions, chemists rely strongly on their knowledge of which reactant acts as nucleophile and which as electrophile to deduce the reaction mechanism using the electron-pushing formalism, whereas students require the reaction mechanism to identify the nucleophile and electrophile. If the mechanism is not available, the students rely stronger on structural components (charges, lone elec-

tron pairs) than on function (donating or accepting electrons).[188] Galloway and her co-workers arrive at similar conclusions after analyzing interviews with first-year students who explained reaction mechanisms by drawing arrows and predicting the product. They find four main categories of utterances: Charges, mapping of electrons, step-wise processes and prior chemical knowledge. While these findings are from a sample of first-year students, misconceptions were shown to be robust as they were still found in fourth-year students.[188] In particular, students would often misapply terms, concepts and algorithms taught in class. Furthermore, the students still struggled with mechanistic thinking. Importantly for this work, the authors write that "[t]here was evidence that the relationship between the relative energies and reactivity of molecules still remains unclear to some students" (p. 129).

Finally, many of the general difficulties associated with the structure of chemical knowledge (see chapter 2) also lead to difficulties around mechanistic reasoning. Specifically, students might rely on individual factors or cause-effect relationships in isolation instead of systematically considering all relevant concepts or they might view chemical heuristics as exact deterministic rules.[6]

Energy

Much of this work is targeted towards first- and second-year undergraduate chemistry, chemical engineering and interdisciplinary science students. At this point in their studies, the students have completed two physics courses as well as a thermodynamics course and multiple general chemistry courses, all of which introduce energy as a salient concept. It was shown that first-year students struggle to integrate and transfer from the initial conceptions of energy they are confronted with in everyday life, physics and biology courses.[189] Although they know that breaking a bond demands energy, they fail to draw inferences from it, such as heat being required to break a bond as opposed to being released.[190] One reason for this failure of transfer between contexts may be found in the way different domains introduce energy. Initially, we get introduced to energy in everyday life as something being stored (for example as sunlight in plants or in food) that then will be released at some point. Later, students encounter energy as a scientific concept in physics lectures as a property of macroscopic objects. The most common example is the gravity-induced conversion from kinetic energy to potential energy. In chemistry, potential energy is a property of the system that depends on the relative positions of the atoms. However, this distinction between the macroscopic, gravitational understanding of potential energy and the sub-microscopic inter-particle distance dependent understanding remains often unclear and hence results in incomplete conceptions.[189] In semi-structured interviews, chemistry and biology undergraduate students

conceptualized potential energy mostly as capability, stored energy or as a measure for stability.[191] Again, while these conceptions are meaningful in specific contexts, their overgeneralization results in incorrect reasoning. Cooper and co-workers[191,192] stressed the importance of introducing energy in an interdisciplinary way and to clearly highlight the connection, similarities and differences between the sub-microscopic and macroscopic conceptualization of energy to prevent potential misconceptions. In terms of the conceptual metaphor concept introduced in chapter 2, the introduction of the concept in many different contexts may lead to several overlapping and potentially contradicting mappings between either prior bodily experiences and colloquial understandings of energy or prior knowledge from previous lectures in physics and biology onto the new domain quantum chemistry.

The ambiguous understanding of energy further becomes apparent in studies of the language used in connection to the concept. Lancor suggested a set of conceptual metaphors in biological, chemical and physical contexts that highlight and neglect different characteristics of energy.[193] For example, *Energy as a Substance that Can be Stored* highlights the possibility for energy transfer, but neglects the need for energy conservation. While the interdisciplinary character of energy has the potential to create difficulties for students, it has been shown that energy in the chemical context, that is in atomic structure, is particularly difficult compared to energy in a biological, environmental or physical context.[194]

The studies presented in chapters 5 and 7 are embedded in a course during which students learned how to calculate the potential energy of relevant molecular structures in a reaction quantum chemically. Based on these energies, they then reasoned about the favored reaction path and consequences in the mesoscopic realm such as the ratio of products. In addition to the incomplete conceptions described above, the technical aspect of computational chemistry adds another layer of complexity to the task at hand. A quantum chemical calculation usually consists of an input file that is fed to a program that returns an output file, from which the relevant energy values are extracted. The calculations themselves remain a black box. One goal of this work is to elaborate on a more intuitive approach to quantum chemical calculations that allows for grounding in prior experience.

Chemical bonding

The misconceptions that students hold of individual bonding models have been researched well with a strong focus on either the different kinds of chemical bonds[23] or hybridization.[25,26]

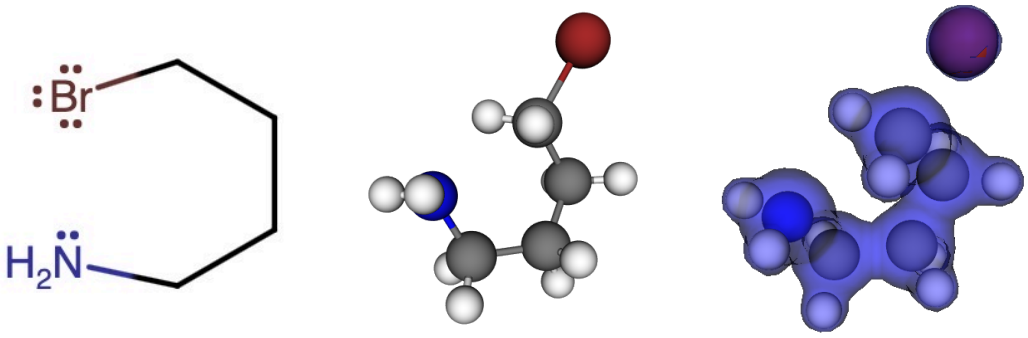


Figure 3.3: Three ways how to visualize a chemical bond. Lewis structures neglect the hydrogen atoms completely and display carbon atoms only as edges, points denote electrons while lines denote electron pairs. The ball-and-stick representation displays the bond as a physical stick that holds the atoms together. The electronic density isosurface marks the area in which the electrons are found with a certain probability.

Many misconceptions arise from either oversimplification or overgeneralization. For example, ionic compounds are often oversimplified by assuming that ionic bonds only exist between alkali metals and halogen atoms[195] or only between single atoms.[196] Furthermore, current teaching methods tend to create the illusion that chemical principles and rules – such as the octet rule – are exact and deterministic over the whole periodic table rather than a heuristic.[197–199] This overgeneralization of the octet rule leads to misconceptions concerning chemical bonding.[197–199]

More generally, Boo found that Grade 12 students conceptualize the chemical bond often as physical entity.[200] Hence, students assumed that as constructing a physical object requires energy, so does the forming of a chemical bond. Similar findings of bonds as entities were also found in university students.[195,201] The reason for this could be manifold. For one, students struggle to grasp the concept of potential energy in a chemical context, because they are introduced to potential energy in multiple contexts such as biology and everyday life or physics lectures.[189] Second, Niebert and Gropengiesser have argued that imperceptible concepts such as potential energy or the chemical bond must be understood *via* a conceptual metaphor.[60] This means that by visualizing chemical bonds as sticks, educators might inadvertently create the metaphor *Chemical Bond is Stick* and map the characteristic of a stick being rigid to the chemical domain (see Fig 3.3 for three ways how to visualize the chemical bond).

When confronted with the quantum chemical understanding of the chemical bond as a less rigid and discrete phenomenon, the students' conception might be challenged without the necessary support to integrate the new information accurately. The traditional curriculum leaps from the chemical bond as a physical construct to one-electron wavefunctions and electronic density. The quantum chemistry concept inventory devel-

oped by Pérez García and co-workers measured the understanding of the chemical bond concept in four items with one specifically focusing on the potential energy.[202] In the administration of the test prior to the quantum chemistry lecture which would correspond quite well to the prior knowledge level of the participants of our studies, many students showed the misconception that "[e]nergy is required to form the bond and that energy is stored within the bond". This answer option stayed popular even after the quantum chemistry lecture. However, decreasing energy was successfully attributed to higher stability by most students in two items. This could mean that the students did not yet associate chemical bonds with stability, an association that potentially could be attributed to the physical entity conception of the bond. Moreover, studies show multiple misconceptions around atomic and molecular orbitals and hybridization.[25,26] As a potential facilitation for this transition, it was suggested to focus on the potential energy diagram of bonding formation.[203,204] In particular, this highlights the discrepancy between a physical entity that requires energy to be built and the chemical bond that lowers the potential energy of the molecular system when being formed.

4

Interactive Quantum Chemistry with SCINE Heron

The Interactive module of SCINE Heron[49] was designed with a strong focus on reducing cognitive load and tailoring towards an audience with low prior knowledge. Simultaneously, the interface was designed to be easily used by experts as well. SCINE Heron offers the graphical interface to multiple modules of the SCINE (Software for Chemical Interaction Networks) project. To increase readability, we will use the term SCINE Heron in place the Interactive module of SCINE Heron. In the following, the design and implementation of SCINE Heron is presented, followed by the results and implications of a usability study conducted with a pilot version of the interface.

4.1 Design

We present a learning environment that allows exploratory learning of basic quantum chemical concepts such as the potential energy hypersurface and its connection to chemical reactivity. For this purpose, we designed specifically for multimodal representations and additionally considered the students prior knowledge and potential misconceptions (see chapter 3). The initial design was primarily a proof of concept of the real-time quantum chemistry framework[11,13,15] with a focus on the specific features needed for learning. Specifically, this included guaranteeing a visual update of the molecular structure at 60 Hz and a force feedback at 1 kHz both based on the gradient of the potential energy hypersurface.[131,132] Concerning the visual representations, the molecular struc-

Table 4.1: Selected load-reduction methods adapted from Mayer and Moreno.[186]

Load-Reducing Method	Implementation in SCINE Heron
Offloading: Move some essential processing from visual channel to *haptic* channel.	Forces which are traditionally displayed as arrows are off-loaded to haptic feedback.
Pretraining: Provide pretraining in names and characteristics of components.	The students performed a set of non-chemical tasks with the haptic device prior to the learning tasks.
Weeding: Eliminate interesting but extraneous material to reduce processing of extraneous material.	Only limited options were shown to the students. This was further improved between the user and the learning study.
Aligning: Place *correlating visual representations close to each other* to reduce need for visual scanning.	The molecular structure and the energy graph are shown close to each other.
Eliminating redundancy: Avoid presenting identical streams of *information*.	Forces are solely represented as haptic feedback.
Synchronizing: Present *haptic feedback* and corresponding animation simultaneously to minimize need to hold representations in memory.	The haptic feedback and energy are displayed simultaneously as the learner is manipulating the molecular system.

Notes. The content was adapted from a focus on auditory and visual channel to haptic and visual channel. Changes are *italicized*.

ture visualization and the energy graph had to be designed meaningfully. By seeing the energy graph together with the movement of the atoms on the computer screen and feeling the resistance against their actions, we hypothesized that the learner may be able to i) ground the concepts activation barrier, potential energy graph and hypersurface and repulsive force in meaningful experiences and ii) connect the two visual representations, molecular structure and energy graph, by creating a haptic bridge.[97]

Our design process was informed by the cognitive theory of multimedia learning by Mayer and Moreno.[99] In this framework, it is argued that information from multimodal sources is processed in distinct channels of the sensory memory and that each channel has its own cognitive load (see section 2.2.2). As this assumption is not modality dependent, we assume that their work, which is focused on the auditory and visual channel, is transferable to our specific case of visual and haptic representations. We applied their suggested load-reduction methods to counteract the potential cognitive overloading of the environment.[186] The considerations are summarized in Table 4.1. We excluded the methods *segmenting* as our environment was designed for self-paced exploratory learning, *signaling* as in a first attempt, we were interested in the students interpretation of the environment without further cues and *individualizing* as this will only be applicable after obtaining user data. The most important design choices include

the *offloading* of the chemical forces onto the haptic modality and the *pretraining*, as we could assume that none of the students have ever navigated a haptic device.

In addition to these initial considerations, we improved the environment iteratively in a usability and multiple learning studies. In the usability study, we were mainly interested in technical feedback and whether the students could interpret and integrate the visual and haptic representations. In the learning studies, we were interested in the effect of the haptic feedback on learning (see chapters 5 to 7). In this chapter, the outcome of the usability study is presented.

4.2 Implementation

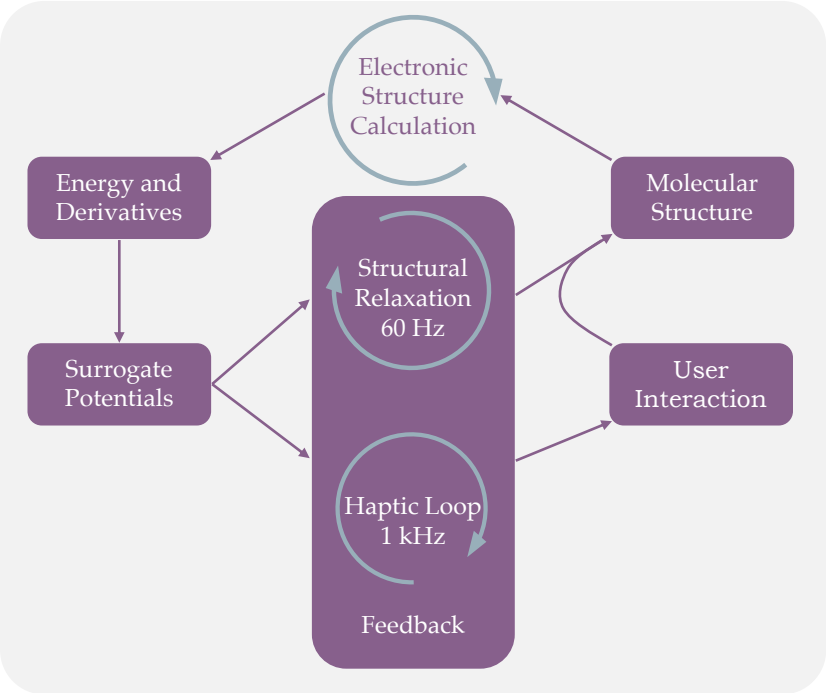


Figure 4.1: Schematic flow in real-time quantum chemistry.[13] The surrogate potentials are second order Taylor series of the electronic energy at the last molecular structure coordinates. Redrawn with permission from A. C. Vaucher, M. P. Haag, M. Reiher, *J. Comput. Chem.* **2016**, 37, 805-812. Copyright 2016 John Wiley & Sons.

The required update rate of 60 Hz for visual feedback and 1 kHz for haptic feedback is guaranteed i) through parameterized quantum chemical methods[21,152] and ii) through a surrogate potential function that serves as proxy between completed calculation cycles.[13] The basic data flow is given in Figure 4.1. Specifically, a surrogate potential V_{sur} is realized as a second-order Taylor series expansion around the atomic coordinates $\boldsymbol{x_0}$ of the last electronic structure calculation. Hence,

$$V_{sur}(\boldsymbol{x}) = V_0 + \boldsymbol{a}^{\boldsymbol{T}}(\boldsymbol{x} - \boldsymbol{x_0}) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x_0})^{\boldsymbol{T}}\boldsymbol{B}(\boldsymbol{x} - \boldsymbol{x_0}) \tag{4.1}$$

with $\boldsymbol{x} = \{x_1, ..., x_m\}$ the atomic positions. V_0 denotes the electronic energy, $\boldsymbol{a}$ its gradient and $\boldsymbol{B}$ the positive definite transform of Hessian $\boldsymbol{H}$ at the reference position $\boldsymbol{x_0}$. That is, $\boldsymbol{a}$ from Equation 4.1 is defined as

$$a_i = \left.\frac{\partial E(\boldsymbol{x})}{\partial x_i}\right|_{\boldsymbol{x}=\boldsymbol{x_0}} \tag{4.2}$$

To avoid unbounded potentials, potentials with negative eigenvalues of $\boldsymbol{H}$ were transformed into potentials with positive eigenvalues $\boldsymbol{B}$ with the procedure described in Equations 4.3 to 4.6.[13]

$$H_{ij} = \left.\frac{\partial^2 E(\boldsymbol{x})}{\partial x_i \partial x_j}\right|_{\boldsymbol{x}=\boldsymbol{x_0}} \tag{4.3}$$

$$\boldsymbol{H_{diag}} = \boldsymbol{U^{-1}HU} \text{ with } H_{diag,ij} = h_{ij} = \begin{cases} 0 & i \neq j \\ b_i & i = j \end{cases} \tag{4.4}$$

$$B_{diag,ij} = \begin{cases} 0 & i \neq j \\ \text{abs}(b_i) & i = j \end{cases} \tag{4.5}$$

$$\boldsymbol{B} = \boldsymbol{U}\boldsymbol{B_{diag}}\boldsymbol{U^{-1}} \tag{4.6}$$

To increase the efficiency of the implementation, atomic Hessians were employed to approximate the full Hessian matrix $\boldsymbol{H}$ in Equation 4.3.[13]

The graphical user interface is Python3 based and builds upon bindings for Qt. Furthermore, it communicates with a Touch™ haptic device by 3D Systems[205] through the open source OpenHaptics library.[206] Any molecular system stored in a `xyz` format can be loaded into the graphical user interface and can hence be manipulated by the learner either with a computer mouse or a haptic device. In real time, the learner can see the electronic energy of the resulting molecular structure in a two-dimensional graph below the molecular structure and simultaneously feel a force feedback that corresponds to the slope of the graph. Specifically, the force feedback is given by

$$\boldsymbol{F_{sur}}(\boldsymbol{x}) = -\nabla V_{sur}(\boldsymbol{x}) = -\boldsymbol{a} - \boldsymbol{B}(\boldsymbol{x} - \boldsymbol{x_0}), \tag{4.7}$$

where $\boldsymbol{F_{sur}}(\boldsymbol{x})$ is a vector of atomic forces which are linearly scaled to match what is perceived as agreeable by humans. This scaling is customizable. A general overview of the development process of the graphical user interface is given in Figure 4.2.

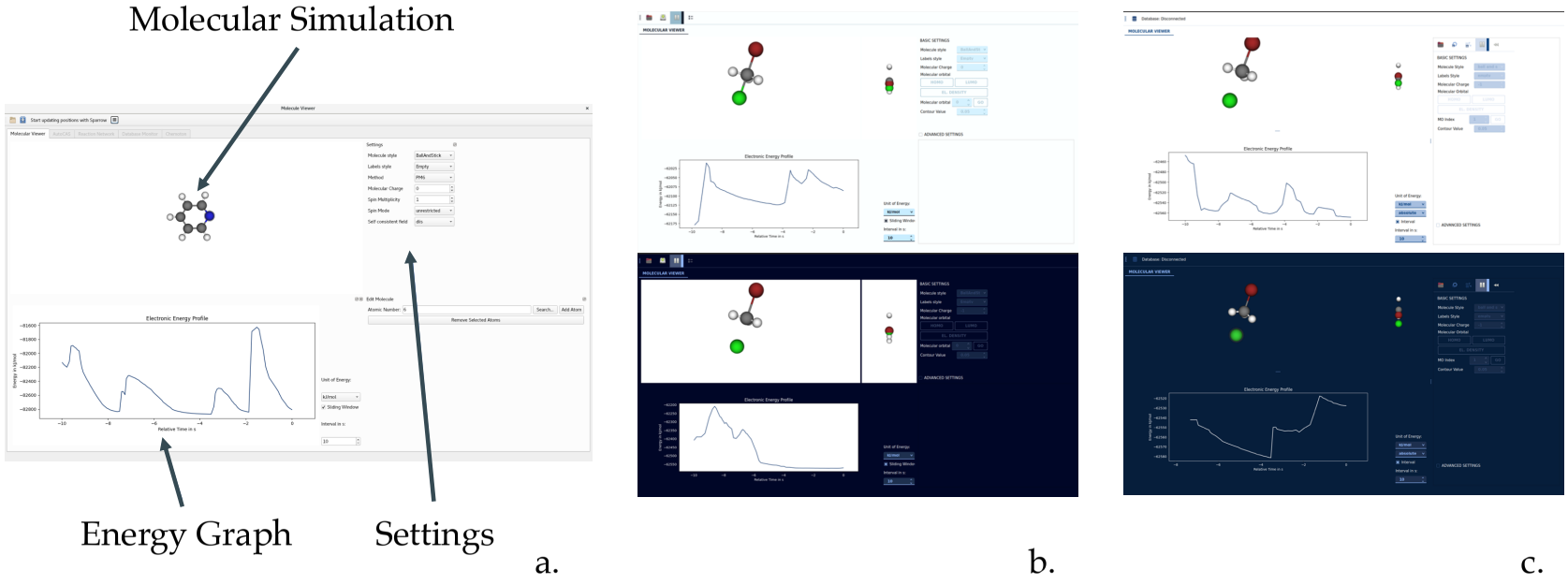


Figure 4.2: Overview of implementation milestones of SCINE Heron. A first implementation was tested by first and third semester students with focus on usability (a.). An improved implementation was applied in two learning studies with second and fourth-semester students (b., see chapter 6 and 5 respectively) and was released after final minor changes (c.[49]).

4.3 Usability study

To study whether the students were able to navigate through the learning environment, we developed three familiarization tasks and three learning tasks. The familiarization tasks focused on basic usability such as drag-and-drop, rotating and zooming. The learning tasks consisted of i) a small molecule on which they could explore the effect that ripping off an atom has on the energy and the force and ii) comparing two nucleophilic substitution reactions, a reaction mechanism that is traditionally taught in the second semester of chemistry or chemistry-related bachelor programs (see Figure 4.3). The guidance was intentionally designed low as we wanted the students to openly explore.

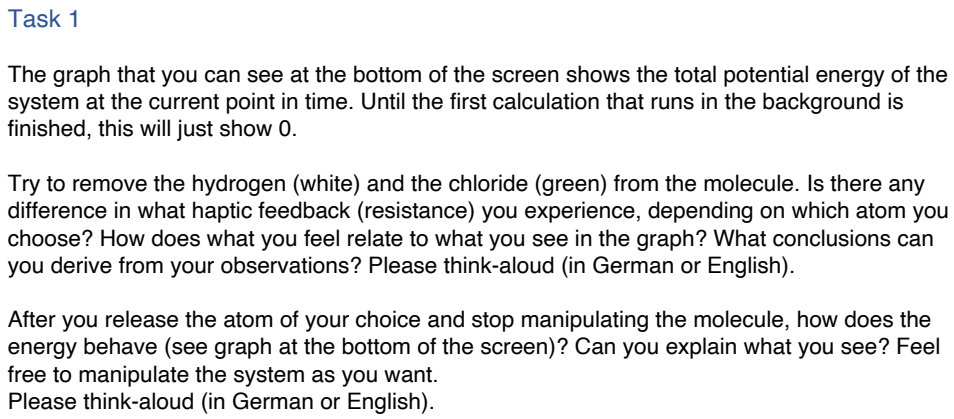


Figure 4.3: Example of exploratory instructions for problem-solving.

4.3.1 Demographics

We recruited first ($n = 9$) and third ($n = 4$) semester students who were enrolled in the bachelor programs chemistry, chemical engineering, biochemistry or interdisciplinary sciences at ETH Zurich (5 females, 8 males, no non-binary people). The students were recruited on a voluntary basis via e-mail and had the chance to win a voucher for the local food market. The age of the students ranged from 17 to 24 (mean $M = 20.23$, standard deviation $SD = 2.09$). All of them reported to be right-handed. The instructions were given in English, questions were answered in German or English, depending on the preference of the student. English proficiency was self-reported to be Conversational ($n = 2$), Fluent ($n = 8$) or Native Speaker ($n = 3$). No student reported to be a beginner in the English language. The third semester students were expected to be aware of the reaction mechanism targeted in the learning environment, while the first semester students were not. The participants of the third semester have received introductory

training in quantum mechanics as part of the introductory physical chemistry course, but not in quantum or computational chemistry.

4.3.2 Protocol

The study session took place in a one-on-one session during the students' free time. The intervention was designed in four parts. First, the students signed the consent form and were evaluated on their spatial ability (mental rotation[207] and visualization[208]). Second, the students were presented with the familiarization tasks. Third, the students completed the learning tasks. The students were asked to think-aloud while performing the tasks, the screen and their voices were recorded. They were free to think-aloud in English or German, the German transcripts were translated to English for reporting. Finally, the students filled out a questionnaire on the usability[209] and immersion[210] of the environment, as well as a learning outcome test (see Appendix A).

4.3.3 Analysis and results

Overall, the written and verbal feedback from the students was positive. The usability was found to be between *OK* and *Good* ($M = 59.0, SD = 4.74$) as defined by Bangor and co-workers.[211] We attribute this score to the lack of familiarity with the device and the intentional lack of guidance provided. As we wanted to give the students as much freedom for exploration as possible, they might have struggled to understand the aim of the environment. The results for the questionnaire on usability and immersion are given in Figure 4.4 and show no alarming patterns. There was no significant effect of the spatial ability on the usability ($t(10) = -0.622$, $p = 0.548$, $d = -4.69$) or immersion score ($t(10) = -0.666$, $p = 0.521$, $d = -0.924$). There was no significant effect of gender on the usability ($t(11) = -0.562$, $p = 0.585$, $d = -17.3$). However, there was a significant large effect of gender on the immersion with the female students generally giving higher scores than the male students ($t(11) = -2.237$, $p = 0.047$, $d = -7.79$). In a collaborative virtual reality study, Schmidt and co-workers found the same effect and attributed it to lesser familiarity and hence less tendency to a habituation effect of the female participants compared to their male colleagues.[212] As we did not find any difference in gaming habits between genders ($t(11) = 0.354$, $p = 0.730$), there has to be another explanation. We tentatively conjecture that as the instructor was female, the female students might have identified with her and were hence inclined to give higher scores.[213]

We conducted an inductive thematic analysis on the think-aloud data.[214,215] We were interested in i) extracting concrete technical feedback and ii) understanding how the students interpreted and, more importantly, connected the three given representations

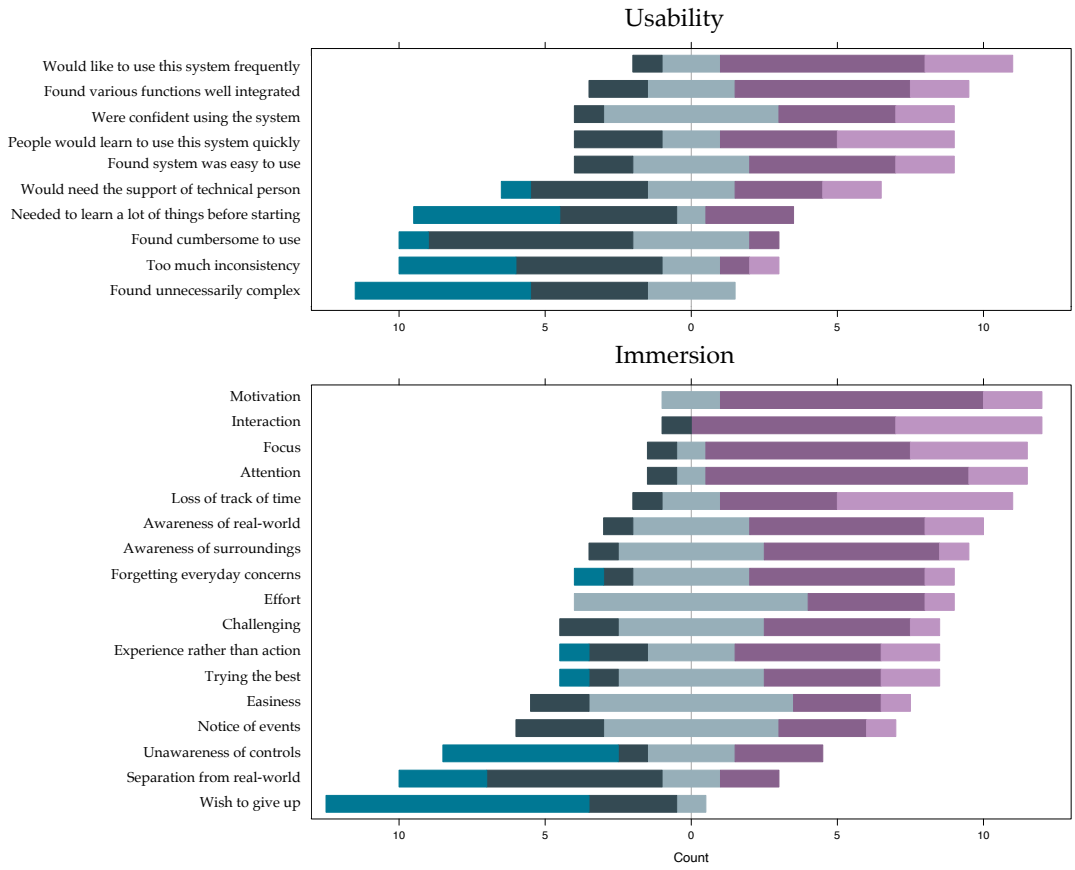


Figure 4.4: Results for usability (five-point Likert scale "Strongly disagree", "Somewhat disagree", "Neither agree nor disagree", "Somewhat agree", "Strongly agree") and immersion (five-point Likert scale "Not at all", "Slightly", "Moderately", "Very", "A lot") from the usability study with the first prototype. For visualization, the items are presented in an abbreviated fashion. They were not changed from the published versions.[209,210]

molecular structure, energy graph and haptic feedback. The think-aloud data were hence coded fine-grained first and the results were then grouped into themes. For ii), we analyzed, whether the students only talked about individual representations or whether they connected two or all of them. The connection had to happen in the same line of reasoning or by building on former comments. A list of all codes from which the following conclusions were derived is given in appendix B. We found that only six out of the thirteen students connected all three representations at least once, indicating that the extraneous load was too high. The following quote from N. Y., who managed to connect all three representations, illustrates what reasoning we were aiming for.

> Okay, the more repulsion there is and the more things I am doing, the bigger the [energy] spikes get. [...] There is more energy in the system and because I have done something, for example if I smash the chloride in there, then I

> am putting, I think, somehow energy into it. And this should then be the energy. (N. Y.)

In this version, all settings had still to be set manually by the instructor and were readily seen by the learner. This resulted in increased extraneous load. For example, one student argued that he will not be able to interpret the environment since he does not know what *spin multiplicity* is. Spin multiplicity is a setting that had to be set by the instructor for the calculations to run but that did not have to be understood in order to be able to draw inferences from the environment.

> [...] So, what is meant here with spin? [instructor answers] So, what is supposed to change when the spin is different... [...] Yes, I just feel, that it is repelling. But I don't know yet, what a spin is. [...] I have done this now, but I do not understand the terminology. (J. A.)

To improve the environment, we hid the distracting settings for future studies, specifically the ones described in chapters 5, 6 and 7. Furthermore, we updated the settings automatically if one setting is changed. For example, if the molecular charge is reset, the spin multiplicity is automatically updated to a valid option.

The most frequent feedback concerned the navigation of the three-dimensional molecular world on a two-dimensional screen. Molecular structures and haptic device movement are both inherently three-dimensional. It is therefore possible to move into the screen in addition to moving on the plane of the screen. This result was obtained by receiving repeated explicit feedback on that matter and by identifying many prolonged pauses where students struggled to pick an atom.

> [...] Added to that, selecting certain atoms with the cursor in three-dimensional space was hard, but I think you would get used to that relatively quickly. (O. F. on written survey)

> It is hard to interpret the depth at the moment. (N. Y.)

To facilitate this navigation, we introduced a visual aid for depth perception of the environment. In this representation, the atoms could all be seen projected onto the dimension perpendicular to the laptop screen. The learner would therefore see a line of atoms as well as the pointer. If the pointer is behind the molecule, the atoms of the molecule are shown below the pointer and vice versa. If the pointer was in the same depth as the molecule, the pointer and atoms are shown overlapping in the visual aid (see Figure 4.2).

With this representation, we avoided incorporating mental rotation tasks as we would have produced, if we simply showed the system from another perspective. We further avoided the implementation of a visual aid into the molecular structure representation such as a grid or shadow. This would have been an acceptable solution for a small

molecular system, but it would have not supported the interaction with large molecular systems such as enzymes.

4.3.4 Limitations

In this chapter, we presented the design process of the interactive haptic quantum chemistry environment SCINE Heron. Version 1.0.0 is available for download.[49] In the usability study, we focused on basic usability issues. Based on these results, we improved the depth perception and the visual settings. The resulting design was introduced in multiple learning studies described in chapters 5 to 7.

We explicitly told the students that they were *testing* an application that might be incorporated into the curriculum in future courses. While this was an effective motivator, it further primed the students to focus on giving feedback and evaluating the environment critically instead of focusing on learning. Similarly, the environment was designed explicitly unscaffolded as we were interested in the perception of the environment. While this resulted in valuable feedback, it also decreased the authenticity of the study. In the study presented in chapter 7, the learning environment was incorporated into a course organically.

With SCINE Heron, we have developed an interactive quantum chemistry environment that can be navigated with a haptic device. In this chapter, we outlined the design process and the program's introduction to an educational setting. Interacting with the world governed by quantum mechanics allows for grounding of this traditionally black box in bodily experience. Students can create meaningful metaphors and can be optimally prepared for instruction.

5

The Effect of Unscaffolded Haptic Feedback on Learning Outcome*

Based on the conceptual framework presented in chapter 2, we applied SCINE Heron (presented in more detail in chapter 4) in a learning study aimed at fourth-semester ETH students. In the following, the methods of this study are presented as well as the results and implications for chemistry education. The precise hypotheses are given in section 2.3. Based on this study, a follow-up study with another cohort of fourth-semester students in the same course one year later was designed that is addressed in chapter 7.

5.1 Methods

5.1.1 Participants

The study was conducted with chemistry, chemical engineering and interdisciplinary science fourth-semester undergraduate students (n = 68, 38% female, 58% male, 1% non-binary/third gender). The participation was part of a one-day course and thus mandatory; however, data were only collected with informed consent. Consent was sought from all. Only one student did not give consent. One more student was removed due to missing data resulting in the final sample size of 68 students. Participants were

*This chapter is reproduced in part with permission from C. H. Müller, M. Reiher, M. Kapur *J. Comput. Assist. Learn.* **In press**.

randomly assigned to one of three groups depending on which week of the semester they signed up to the course. The students have learned the reaction mechanism of a nucleophilic substitution reactions approximately one year prior to the study and they were further familiar with simple two-dimensional energy graphs that statically depict a reaction as traditionally found in text books. They did not yet explicitly learn how force is related to the energy and they were not yet aware of the inherent multidimensionality of the potential energy hypersurface according to the curricula of their programs. Moreover, as mentioned in previous sections, they potentially activated prior conceptions of energy in a physics context. Table 5.1 gives an overview of the baseline statistics. The study was approved by the ETH ethics committee and the study design, analysis plan, anonymized data as well as analysis scripts are available on OSF [†].

5.1.2 Material

The learning environment in SCINE Heron

The learning environment in SCINE Heron is an open-source graphical quantum chemistry environment that allows the calculation of electronic energies of and interact with chemical systems in real-time.[49] A detailed account of the design process and results to a usability study are presented in chapter 4. Importantly to the study, the environment can either be controlled with a Touch^{TM} haptic device or a computer mouse (see Figure 5.1). The learner selects an atom and picks it with the haptic device. The rest of the molecular system reacts according to these actions by continuously running energy minimization. The design process was informed by the theory for multimodal learning, which emphasizes that the capacity for cognitive load is limited and that there is a chance of overloading when introducing a multimodal learning environment.[186,216,217] Informed by the guidelines from Mayer and Moreno[186] on how to reduce cognitive load, we incorporated pretraining to make the students familiar with the haptic device prior to problem-solving. Moreover, we eliminated redundancy by presenting atomic movement in the simulation, energy change in the energy graph and force as repulsion and attraction of the haptic device. The latter could be visualized by force vectors on each atom. Here, we are offloading this representation onto the haptic modality. However, it should be noted that the force can still be derived visually by observing how the spectator atoms react to the manipulations of the learner on a selected atom. If the spectator atoms follow the chosen atom closely so that the learner has to move the haptic pointer quickly to break the bond, the withdrawing force is larger than if the learner can easily remove the atom from the remaining molecule. However, this overlap in information

[†] https://osf.io/t9wne/?view_only=3c6fdb25cf6947688f56698239f6b2d4, accessed 2023-08-04

Table 5.1: Baseline characteristics for the participants. $N = 68$ ($n_{control} = 22$, $n_{comp} = 21$, $n_{exp} = 25$). Participants were on average 20.9 years old ($SD = 1.6$) and the groups did not differ in age. The control group completed the problem-solving tasks with a computer mouse, the comparison group with a haptic device with deactivated haptic feedback and the experimental group with a haptic device with activated haptic feedback. All characteristics are self-reported.

Baseline characteristic	Control		Comparison		Experimental	
	n	%	n	%	n	%
Gender						
Female	9	40.9	6	28.6	11	44.0
Male	12	54.5	15	71.4	12	48.0
Non-binary / third gender	0	0.0	0	0.0	1	4.0
Prefer not to say	1	4.5	0	0.0	1	4.0
English proficiency						
Conversational	5	22.7	4	19.0	5	20.0
Fluent	17	77.3	17	81.0	16	64.0
Native Speaker	0	0.0	0	0.0	4	16.0
Gaming						
Less than once a year	7	31.8	5	23.8	16	64.0
Once a month	6	27.3	5	23.8	4	16.0
Once a week	1	4.5	6	28.6	2	8.0
More than once a week	8	36.4	5	23.8	3	12.0
Handedness						
Left	2	9.1	1	4.8	2	8.0
Right	18	81.8	19	90.5	23	92.0
Both	2	9.1	1	4.8	0	0.0

can also facilitate connection between the representations.[98] We discuss the extent to which this creates potentially unnecessary redundancy. Finally, the haptic feedback was synchronized with the movements and changes in energy as it corresponded to the derivative of energy which itself depends on the movements of the atoms. The haptic feedback can therefore alternatively be described as a haptic bridge[97] between the two visual representations, the energy graph and the molecular simulation. Specifically, learners see that, as they move an atom, they feel a resistance or attraction and they see that the energy graph changes.

Furthermore, we implement the grounded and embodied learning principles[3] in that the students can ground the concept of force in a chemical context to the perceptual process

Figure 5.1: Re-enactment of the learning environment set up. The control group received the same visual environment with a computer mouse and without the depth perception aid, as the computer mouse can only move in two dimensions.

of experiencing repulsive force. Moreover, modal priming allows for mental simulations of sensorimotor experiences to inform future cognition.[85] Modal priming describes how sensorimotor states can potentially activate abstract concepts. In this specific case, the sensorimotor experience of repulsion upon steep ascend of the energy graph can inform the students' later reasoning about chemical reactivity and specifically activation energy.

This learning environment shows high resemblance to the software ELI Chem by Zohar and Levy.[20] While with SCINE Heron, students could similarly move an atom toward or away from other atom(s) and simultaneously feel the repulsion or attraction that would operate on the selected atom, the rest of the molecular system additionally visually adapted. Moreover, in our set-up, the movements occurred in three dimensions as opposed to two in the case of ELI Chem. Finally, in our study, the students were asked to perform reactions as opposed to explore solely the bonding process between two atoms.

Problem-solving Tasks

The problem-solving phase took place in four parts. First, the students were introduced to the learning environment through a short written tutorial of which a part can be

seen in Figure 5.2, left. Specifically, they learned how to select and drag an atom, how to zoom or rotate and how the force feels like for the students who received haptic feedback. Again, the force corresponded to the gradient of the energy at the selected atom. In other words, if the selected atom is released, the gradient indicates where the atom would spontaneously move to. The students further trained their ability to control the three-dimensional environment by pulling an atom around another molecule or by guiding an atom through a carbon tube. The next two parts were then presented as open problems aiming to prepare the students to learn about i) how energy change is related to force and hence molecular structure change and ii) what this implies for chemical reactions. As the instructions were equal for all groups, the force and its correlation to the slope of the energy graph were not explicitly mentioned. The instructions solely consisted of the task descriptions, as well as open-ended questions.

First, the students received a single molecule that they could manipulate at will. This potentially led to the understanding of force being related to movement of atoms and of the change in energy upon structural change. Specifically, the students were asked to contrast and compare between removing a hydrogen atom and a chloride atom. In the next part, the students received two reactants of a simple nucleophilic substitution (see Figure 5.2, right). They were asked to perform the reaction in two different approaches and to compare and differentiate between them. Furthermore, they could explore the influence of a change in molecular charge on the outcome. Finally, in the third part, they received again two different reactants of a nucleophilic substitution. Hence, the reaction mechanism was the same, but the molecules were different, allowing the students to test previous hypotheses and to abstract from their previous observations. The students who received haptic feedback could build their hypothesis on the force that they had to apply in order to perform the reaction while the other students had to rely on the energy graph and the spontaneous movements of the unpicked atoms. Throughout the two problems, the students were asked open-ended questions that they responded to in writing (see Figure 5.2, right). For example, one question was “How does the potential energy change? Do you notice a relation between this change and your movements?”. The progression of these answers could indicate the learning trajectories of the students.

Instruction Phase

The instruction phase consisted of a video presentation. The visuals included a screen recording of the learning environment that the students previously interacted with. The instructor explained to them, how the force specifically relates to the change in energy. Moreover, the frontier molecular orbitals were shown to give the students an idea about

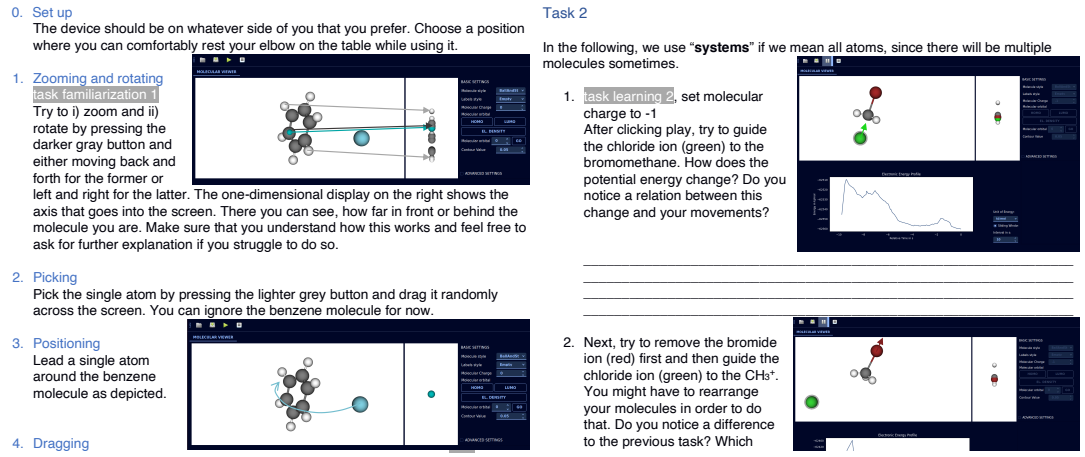


Figure 5.2: Excerpt of familiarization tasks (left) and the second problem-solving task (right) of the problem-solving phase. The written description was accompanied by screenshots of the environment for the students to follow.

how they can further inform reaction outcome. This was mainly needed for later points in the class that was not part of the study. The video was four minutes long.

Quantitative Measures

The independent variable was the mode of manipulation that the students employed during the problem-solving phase of the study. While receiving the same visual feedback, they either controlled the learning environment on the computer screen with a computer mouse (control), with a haptic device without receiving haptic feedback (comparison) or with a haptic device receiving haptic feedback corresponding to the energy change in the environment (experimental).

The dependent variable was performance which was measured through an in-house developed posttest. The questionnaire was designed in four parts, potential energy surface, force, reactivity and S_N2 as isomorphic example. In particular, we were interested in their understanding of the role of energy extrema and saddle points, slope and their connection to molecular structure changes such as stretching and forming of bonds. The full questionnaire including the learning objectives is available in Appendix A. The multiple-choice posttest was validated by two chemists and refined twice, once by online distribution to students and once based on the results in a previous usability study for SCINE Heron (see chapter 4). The test measures multiple interconnected concepts such as change of potential energy related to molecular structure, change of energy related to force, energy difference related to reactivity or energy extrema as critical points during reaction.

Moreover, cognitive and affective mechanisms which have been shown to be associated with problem-solving were measured through published questionnaires[19] (see Table 5.2 for an overview). In the following, the references refer to the questionnaires that were used in total or in part. Specifically, we measured positive and negative affect,[218] state curiosity[219] and knowledge gap awareness.[220] To study the cognitive load associated with controlling the haptic device, we further measured the cognitive load that did not directly contribute to learning (extraneous load[220]). The item "I found it hard to find all the information necessary to solve the task" was removed as it rather measured the task difficulty than the extraneous load and hence decreased the reliability significantly. Finally, we measured the agency and motor control as well as the response to external stimuli through the corresponding items from the avatar embodiment questionnaire by Gonzalez-Franco and Peck.[221] We removed the items "The movements of the pointer were caused by my movements" and "I felt as if the pointer was moving by itself" due to poor reliability. Moreover, as the pointer did move by itself, for the experimental condition, the items do not seem applicable.

The average over the students' general chemistry grades of the first year were taken as a proxy for prior content knowledge. These included the introductory organic chemistry, introductory inorganic chemistry and introductory physical chemistry grade.

Learners might find it challenging to translate between a two-dimensional computer screen and three-dimensional movements, depending on their spatial ability. We were hoping that the familiarization tasks at the beginning of the problem-solving phase would reduce any differences in that regard and to guarantee that no discrimination based on spatial ability is introduced, we measured it through the measures for mental rotation[207] and visualization.[208]

Finally, their English proficiency (self-reported), gender, age, gaming proficiency and handedness was measured.

Table 5.2: Overview of learning mechanism and content knowledge measures.

Construct	Example item	Type	Total Omega	Source
Extraneous load	It was easy for me to distinguish important from unimportant information.	Likert scale	0.66	Original items from Glogger-Frey et al.[220]
Positive affect	Interested	Likert scale	0.87	Original items from Watson et al.[218]
Knowledge gap awareness	My knowledge was insufficient to carry out the tasks.	Likert scale	0.88	Original items from Gloger-Frey et al.[220]
State curiosity	I want to know more.	Likert scale	0.96	Original items from Naylor.[219]
Avatar embodiment	I felt like I could control the pointer as if it was my own hand.	Likert scale	0.87	Adapted agency and motor control & external stimuli response items from Gonzalez-France & Peck.[221]
Content knowledge	What do individual points on a potential energy graph correspond to? a. Different spatial arrangements of atoms. b. Different levels of excitation. c. Different vibrational frequencies. d. Different atomic velocities.	Single-choice	0.62	Developed in-house.

Qualitative Measures

The learning trajectory was captured through open-ended questions embedded in the instructions. Examples of such questions are "What differences or similarities did you notice?", "What changed?" or "Are your conclusions coherent with what you observe?". They were answered in writing on paper.

5.1.3 Design

PS-I prepares the students for instruction by activating their relevant prior knowledge, creating awareness of knowledge gaps and by guiding them to recognize deep features of the problem.[9,16,17,19] Typically, students attempt to solve a problem that requires the application of a yet to-be-learned concept and, usually, they are not fully successful in this respect.[18] In the failing, the students realize the gaps in their knowledge by identifying the deep features of the problem and the information that they would need to overcome or understand them. In the instruction phase, they are then presented with the solution. Now that the students have a better understanding of the problem, they can better appreciate the solution and understand it more deeply.

In this study, we took an exploratory approach to invoke these learning mechanisms. In the first problem-solving task, the students were first made aware of the connection between the electronic energy and the molecular structure by manipulating a single molecule. In the second and third task, the students explored two nucleophilic substitution reactions, meaning two reactions that follow the same mechanism but that involve different reactants. In open-ended questions, they were asked to connect the representations and to compare the different approaches with the goal of identifying deep features of this specific reaction mechanism and of chemical reactivity in general – mechanism elucidation as minimal energy problem, the role of molecular and partial charge and frontier molecular orbital theory. After completion of the problem-solving tasks, the instructional video showed a screen recording of the learning environment and explained, how the spatial arrangement of the atoms is corresponding to the energy and how the reaction that they observed can potentially be explained through the frontier molecular orbitals.

5.1.4 Procedure

The experimental study followed a preparation for future learning design.[9,16,19] Prior to the course, the students filled out a demographic questionnaire. In the course, they first filled out the spatial ability tests. Then, they started by getting to know the environment with simple manipulation tasks such as rotating, zooming or dragging and dropping of

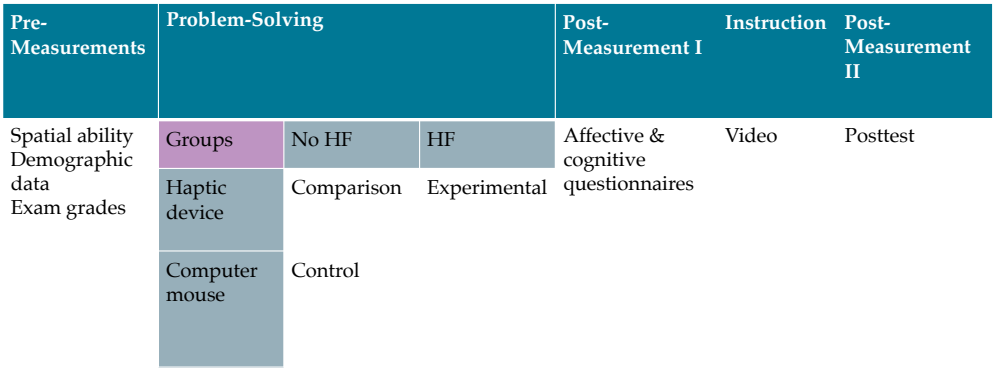


Figure 5.3: PS-I design. HF denotes the haptic feedback. Post-measurement I occurred immediately after the PS phase and post-measurement II immediately after the instructional video.

an atom around another molecule. After self-proclaimed proficiency, they moved on to the problem-solving phase in which they were asked to interact with three chemical systems. The first one involved one single molecule with which the students were able to explore the energy change and potentially associated haptic feedback resulting from the interaction with the molecule. Specifically, they were asked to remove a chloride ion and a hydrogen atom and compare the outcomes. The other two systems were nucleophilic substitution reactions. In addition to exploring different reaction paths, they could compare the reaction with different molecular charges. In the last task, they could additionally look at the electron density and molecular orbitals as well as at the atomic charges. During the problem-solving phase, they answered open-ended questions in writing directly onto their instructions. After finishing all the tasks, they were asked to fill out the questionnaire measuring the cognitive and affective variables. The instruction phase followed in the form of an instructional video. Finally, they filled out a posttest that measured their understanding of the target concepts (reaction energy, activation energy and reaction coordinate), as well as near transfer (potential energy surface). The procedure took approximately 90 min. An overview is given Figure 5.3.

5.1.5 Analysis

Quantitative analysis

Hypothesis 1 was tested in an ANCOVA with dependent variable posttest performance, independent variable group and covariate prior content knowledge. To study the novelty effect, we conducted a planned contrast between the control group and the two groups involving the haptic device. Finally, haptic feedback was isolated by comparing the experimental and comparison group.

Hypothesis 2 was tested in a parallel mediator analysis with the PROCESS macro in R 2022.12.0.[222] We included learning mechanisms that are known for being salient in PS-I, knowledge gap awareness, positive affect and state curiosity as mediators in the model as well as prior content knowledge as covariate.

Qualitative analysis

The quantitative results resulted in the need to look at qualitative data post hoc. Hence, the qualitative analysis is inherently exploratory for the purpose of developing tentative explanations and descriptions of how the students engaged with the learning environment. For the analysis of the open-ended questions posed during the problem-solving phase and answered on paper, we followed a thematic analysis approach as described in Braun and Clarke.[214,215] We adopted an experiential realist stance. This implies that we assumed that knowledge and here specifically the students' knowledge emerges from their active experiences. Consequently, knowledge is constructed rather than discovered.[223] It should be noted though that as I was a novice to qualitative research and have a background in chemistry, my initial postpositivist position (perspective of probabilistic outcome, meaning that there is a truth to be discovered by the researcher[224]) shifted towards a realist perspective only during the analysis and hence my personal learning process. This also explains my initial development of a coding book from which then the themes were developed.

Due to the nature of the research questions, we were interested in how students perceived and interpreted the force, both visually and haptically. Moreover, we were interested in what models were frequently activated during the experience generally, as well as with regard to the calculation. Finally, we were interested in what effect the spontaneous movement of the atoms had on the students as this may be interpreted as another representation of force. Generally, we were looking for explanatory accounts for the quantitative results, precisely since the latter appeared to contradict our conceptual framework. Therefore, we were following both deductive and inductive approaches. The analysis was limited to the two haptic device groups as we were interested in the effect of haptic feedback specifically. Initial codes were generated fine-grained until convergence. Second, we went through the individual codes to see, if the coding made sense for the selected context or if more context was needed. Finally, we compared the groups with respect to the scope and nature of activated models, programming related codes and indication of focus.

5.2 Results

Table 5.3: Overview of descriptive statistics of the continuous variables. The spatial ability and prior knowledge are incoming characteristics, the others outcome variables. All values are given in percent. $\tilde{\mathbf{x}}$ denotes the median, $\bar{\mathbf{x}}$ the mean and **SD** the standard deviation.

Variable	Group	n	Min	Max	$\tilde{\mathbf{x}}$	$\bar{\mathbf{x}}$	SD
Spatial ability	Control	22	19.2	85.0	53.5	57.3	16.1
	Comp.	21	27.5	70.8	47.1	49.4	12.5
	Exp.	25	12.5	81.2	53.8	52.2	16.9
	all	68	12.5	85.0	52.5	53.0	15.5
Prior knowledge	Control	21	59.7	100.0	70.8	74.8	11.0
	Comp.	20	25.0	91.7	70.1	68.6	18.6
	Exp.	25	63.9	98.6	77.8	78.7	10.2
	all	66	25.0	100.0	73.6	74.4	13.9
Extraneous load	Control	22	33.3	83.3	62.5	64.4	10.7
	Comp.	21	33.3	87.5	62.5	62.3	12.9
	Exp.	25	45.8	75.0	58.3	61.8	9.2
	all	68	33.3	87.5	62.5	62.8	10.8
State curiosity	Control	22	28.8	93.8	58.8	62.3	17.4
	Comp.	21	51.2	93.8	70.0	70.2	11.4
	Exp.	25	25.0	96.2	71.2	71.2	18.1
	all	68	25.0	96.2	68.8	68.0	16.3
Knowledge gap awareness	Control	22	20.0	93.3	36.7	42.0	18.6
	Comp.	21	23.3	73.3	46.7	48.6	13.3
	Exp.	25	26.7	86.7	46.7	48.1	17.2
	all	68	20.0	93.3	43.3	46.3	16.6
Positive affect	Control	22	24.0	78.0	43.0	48.8	17.2
	Comp.	21	30.0	74.0	58.0	55.0	13.0
	Exp.	25	22.0	86.0	54.0	54.8	16.6
	all	68	22.0	86.0	53.0	52.9	15.8
Embodiment	Control	22	31.7	79.4	48.4	50.6	14.0
	Comp.	21	28.6	90.5	49.2	48.9	14.0
	Exp.	25	49.2	87.3	68.3	68.3	10.5
	all	68	28.6	90.5	55.6	56.6	15.5
Posttest	Control	22	33.3	91.7	66.7	63.6	17.4
	Comp.	21	33.3	91.7	58.3	65.5	17.3
	Exp.	25	33.3	91.7	58.3	62.3	15.8
	all	68	33.3	91.7	66.7	63.7	16.6

5.2.1 Performance Analysis

Prior Knowledge Proxy

The average over three first-year general chemistry grades (physical, inorganic and organic chemistry) were taken as proxy for prior content knowledge and were included to the model as covariate. The prior content knowledge was significantly non-normally

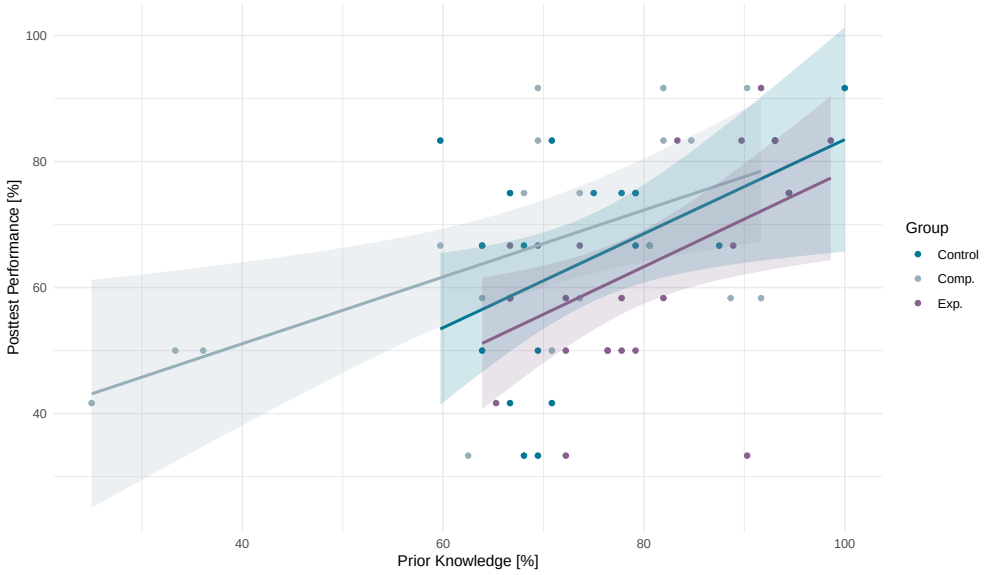


Figure 5.4: Scatter plot comparing the covariate prior knowledge and the dependent variable performance. Removing the three data points that visually appear to be outliers (but statistically are not) does not change the outcome. As all groups exhibit comparable slopes, prior knowledge is a suitable covariate.

distributed overall ($W = 0.91$, $p < 0.001$) as well as in the comparison ($W = 0.89$, $p = 0.022$) and control group ($W = 0.90$, $p = 0.034$) which is why the findings were reproduced in a robust ANCOVA for two groups and one covariate. The assumption of homogeneity of regression slopes for prior content knowledge ($F(2, 60) = 0.32$, $p = 0.728$), the assumption of homogeneity of variance and the assumption of independence of prior content knowledge and treatment effect ($F(2, 63) = 3.12$, $p = 0.051$) were met. An overview of the descriptive statistics of the outcome variables is given in Table 5.3. Furthermore, the data are visually presented in Figure 5.4.

Analysis of Covariance

The multiple-choice posttest items were either coded correct or incorrect. We conducted a standard ANCOVA with the posttest performance as outcome and the groups as predictor with the prior content knowledge as covariate. The Shapiro Wilk test was not significant ($p > 0.05$), indicating normality of residuals. The Levene's test was not significant ($p > 0.05$), indicating homogeneity of the residual variances for all groups. There were no outliers in the data, as assessed by no cases with standardized residuals greater than 3 in absolute value.

We did find a medium effect of spatial ability on the learning outcome ($F(1, 62) = 6.26$, $p = 0.015$, $\eta^2 = 0.091$), but no interaction effect with the groups on learning outcome

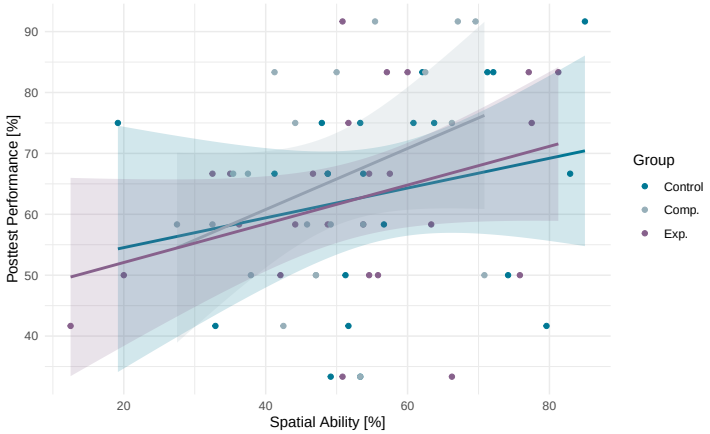


Figure 5.5: Spatial ability had a significant effect on learning outcome that was independent of the condition.

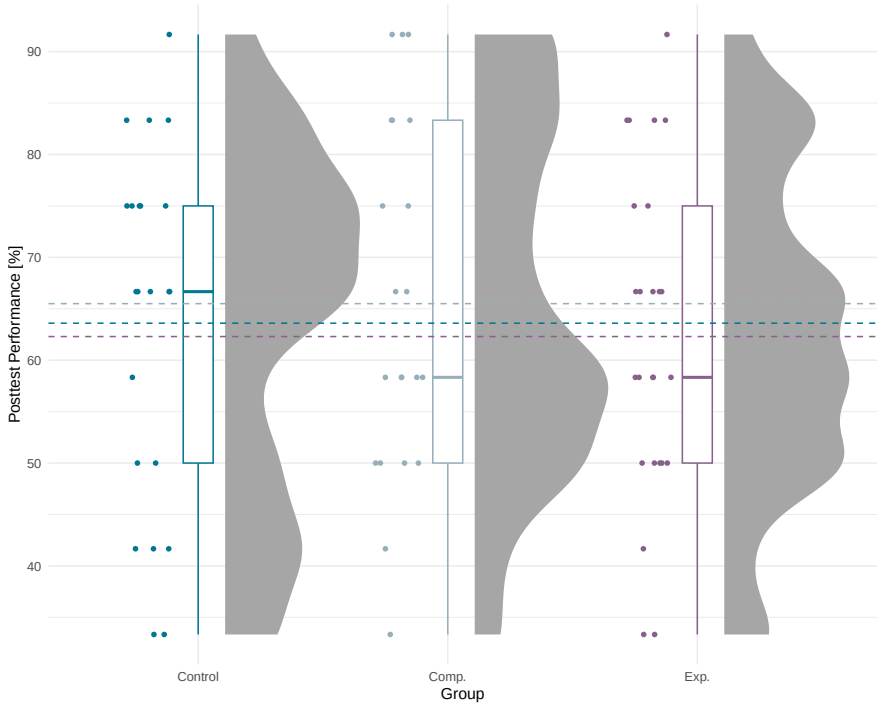


Figure 5.6: Distributions of posttest performance for the individual groups. The means are indicated with dashed lines.

($F(2, 62) = 0.25$, $p = 0.782$, $\eta^2 = 0.0071$, see Figure 5.5). Prior content knowledge had a medium effect on the learning outcome, $t(62) = 4.61$, $p < 0.001$, $d = 0.70$ with a correlation of $r(64) = 0.45$, $p < 0.001$. Furthermore, a planned contrast analysis revealed that receiving haptic feedback had a small negative effect on learning outcome, $t(62) = -2.24$, $p = 0.029$, $d = 0.19$. The differences in learning outcome per group are visualized in Figure 5.6. There was no significant difference with respect to the tool of manipulation (computer mouse or haptic device), $t(62) = 0.09$, $p = 0.932$, $d = -0.01$, indicating that there was no significant novelty effect of the device alone. As the data did not meet all assumptions required by a standard ANCOVA, we replicated the found difference between the comparison and experimental group with a robust method.[225] The robust ANCOVA compares trimmed means at different points along the covariate prior knowledge at which the relationship between the covariate and the outcome are the same. All of the five design points sampled by the method showed a significant or marginal relationship of similar magnitude between the group and the posttest (see Table 5.4).

Table 5.4: Robust ANCOVA results from the WRS package[225] comparing the experimental to the comparison group. $\cdot < .1$, $* < .05$. Prior content knowledge is included as covariate. n_i denote the number of subjects of group i at evaluation point. Further reported are the differences between the trimmed means including its standard error SE and confidence interval CI.

Prior knowledge [%]	n_{comp}	n_{exp}	Trimmed mean diff.	SE	CI	p-value
66.67	14	13	-12.78	5.27	[-28.39, 2.83]	0.030∗
69.44	16	14	-14.17	5.22	[-29.22, 0.88]	0.015∗
73.61	18	15	-11.81	5.61	[-28.62, 5.01]	0.057·
77.78	18	16	-10.97	5.47	[-27.04, 5.09]	0.064·
83.33	18	13	-13.19	7.43	[-34.42, 8.03]	0.093·

Mediator Moderator Analysis

Finally, we aimed to conduct a mediator moderator analysis to gain further insight into the salient learning mechanisms for learning from haptic feedback. The analysis was not significant for all learning mechanisms state curiosity, knowledge gap awareness, affect and extraneous load, neither was the interaction between the moderator spatial ability and the group on the extraneous load for both contrasts hapic versus non-haptic and device versus computer mouse. The only significant result is associated to the avatar embodiment which served as a manipulation check and, as in the contrast analysis, the direct effect of the force on the learning outcome was significant. The detailed results

are given in Figure 5.7. This was to be expected, as the groups themselves differed only marginally in the learning outcome. Furthermore, the study was underpowered for an analysis including eight predictors ($1 - \beta = 0.684$ with $\alpha = 0.05$).

5.2.2 Thematic analysis

A summarized version of the final coding book is given in Table 5.5 and the full list of codes is given in appendix B.

During the analysis process, we found that simply gathering and collecting codes is not sufficient to explain the quantitative results. Therefore, we further reviewed the final topic summaries and generated five themes that are further discussed in this section. Three of these, *energy is proportional to the force*, *binary understanding of chemical bonds* and *energy level vs. energy well* are based on codes dealing with the models that the students employed. In them, we try to retrace the reasoning of the students to prior conceptions and how they might have been activated by the external representations in the learning environment. The theme *atomic movement* was generated from codes in the Focus topic summary. We try to explain how the students interpreted and connected the different external representations and the role that the atomic movement was playing in it. Finally, the theme *lack of trust* examines confidence of the students in the program. The examples are given as literal as possible including mistakes with German quotes being translated to English.

The students generally employed two main models when talking about the electronic energy, the *energy level* and the *energy well*, regardless of the group. The former was found mainly in the context of describing the energy graph. For example, C. G. writes that "[t]he energy returns to the original state, if you don't move the atoms too far away." or B. B. mentions when comparing energy values that "[t]he energy levels out, if the atom is reattached to the molecule the energy is the same as the baseline energy, if the atom is removed it is higher". By contrast, when students mentioned extrema of a continuous surface, we denoted this as the energy well model, as for example in "If the atom is let go, the bond breaks and the potential energy goes back to its minimum." (M. G.). These models were not activated discretely but were both explicitly mentioned by nine students. Only six mentioned solely the energy level model and four mentioned solely the energy well model.

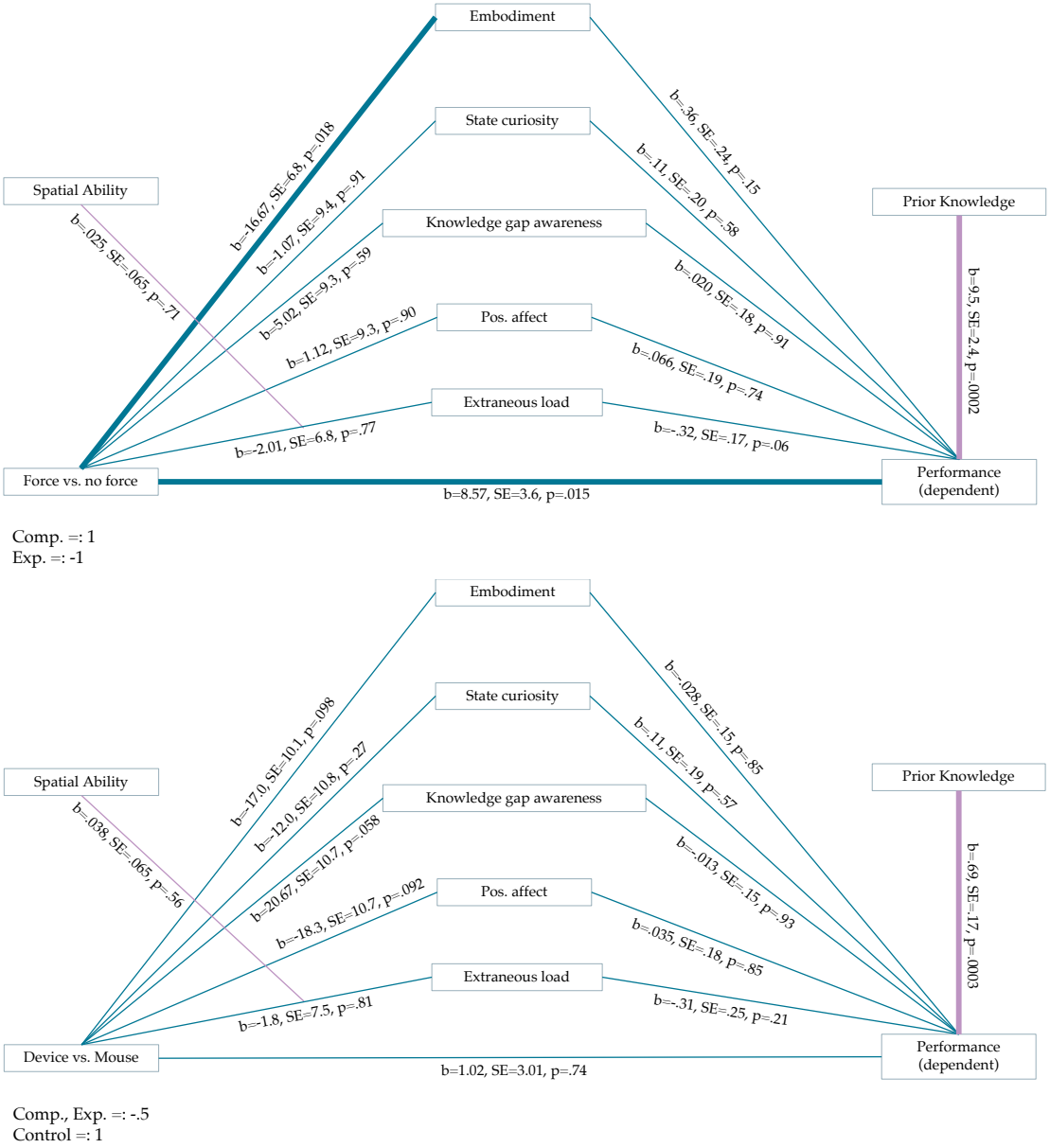


Figure 5.7: Effect of learning mechanisms on learning with or without haptic feedback (top) or with a haptic device or with the computer mouse (bottom). Significant relationships are indicated with bolder lines. Prior knowledge was a significant predictor for performance in both analyses. The main effect of the haptic feedback was also again significant. Finally, avatar embodiment was higher for students who received haptic feedback.

Table 5.5: Final coding book for thematic analysis of written answers to the open-ended questions posed during the PS phase.

Topic Summary	Codes	Description	Example (Code)	Cases	References
Focus	- Connect graph - atomic movement - Connect graph - haptic - Focus graph - Focus haptic - focus molecule	Codes that describe on which external representation(s) the students focused.	There is a big increase of potential energy and there is a big repulsive feeling. (connect graph-haptic)	45	288
Model	- Bond binary - Charge - Energy - Force	Codes that identify scientific models that the students used (not necessarily accurate ones).	Energy rises when removing both the hydrogen and the chloride. (model gradual E change)	43	155
Program(ing)	- Aware of moving atoms - Unaware of moving atoms - Trust - Usability	Codes that identify notes by the students on the program or the programming incl. technical issues or usability feedback.	The iodine charge seems rather high and has less effect on neighboring atoms than expected. (program trust)	27	50

We further identified a mainly *binary understanding of chemical bonds*. The students attributed certain characteristics or changes in energy and forces to the point of formation or breaking of the bond. For example, S. M. wrote that "[a]s soon as the bond is formed, the energy is higher than before" or V. D. wrote that "[i]t went down again when the chloride bound to the molecule". M. M. even expected a change in properties at the point of bond breaking: "E[nergy] rises the closer I move Cl to C even before + after bonding".

Concerning the representations, we found that students who did not receive haptic feedback extracted similar information from the *atomic movements* as the other group from the haptic feedback. Faster atomic movement was interpreted as more force and was connected to a change in the energy graph. In addition, the students did not appear to be able to perceive and integrate all three dynamic representations, the energy graph, the atomic movements and the force. Students who received haptic feedback tended to either focus on the molecular simulation or on the haptic feedback. In some cases, the atomic movement was even explicitly not perceived. For example, A. T. writes that "[t]he chloride ion then *almost* moves towards the CH_3^+ on its own, what also correlates with a decreasing energy" and H. L. does not appear to be sure when they write that "maybe, the atoms make new bonds on their own". M. G. writes that they "would expect the energy of the system to stay at the same level if nothing is manipulated, as is to be expected", when in reality, the energy did definitely decrease once they let go of the picked atom due to the continuous structural relaxation.

Regarding the haptic feedback, we found that it did not necessarily support the learning of the concept force as change of energy. In many cases, it simply supported the misconception that the *energy is proportional to the force* or that energy and force can be employed synonymously. For example, A. C. writes that "[...]. This means, the stronger the resistance, the higher the energy" and J. S. concludes that "[w]hen I remove bromide first, the energy is much larger, I have to pull harder". In the case of B. B., this conception was challenged, as they found that "[t]he chloride seems to give more resistance, but on the graph, the energy peaks for both are about the same height". However, in many cases, it was not clear whether force was associated to a change in energy or energy itself, as the students would often describe that the energy rose and they could simultaneously feel a resistance.

Finally, we found some indications of *lack of trust* in the environment. For example, F. P. wrote that "the partial charge on I[odine] should be smaller, in this case the partial charge on N is smaller than I[odine]". Here, it was not trivial to distinguish between inability to explain an event and distrust in the environment. The example given is the only explicit unambiguous criticism found.

5.3 Discussion

In the presented mixed-methods PS-I study, we tried to reproduce the positive effect on learning outcome of haptic feedback. Hence, we compared three groups who completed the same problem-solving tasks with different manipulation tools. The control group employed a computer mouse, the comparison group a haptic device with deactivated haptic feedback and the experimental group a haptic device including feedback. We hypothesized that posttest performance will be highest for the students receiving haptic feedback (experimental), followed by students who employed the computer mouse (control) and last by students who employed the haptic device but did not receive haptic feedback (comparison) due to increased extraneous load.

Our findings suggest the very opposite, the haptic feedback showed hindered learning as indicated by significantly lower posttest performance in the group who received haptic feedback. Furthermore, we found that the haptic device alone did not have an effect on learning, indicating that the pretraining of the device was sufficient to limit the additional cognitive resources needed to control this unfamiliar device. This is supported by the absence of a significant interaction effect between spatial ability and treatment on the learning outcome.

Previous studies which showed a lack of significant results in favor of the haptic feedback attributed this to the increased cognitive load.[10,183] However, we did not find a difference in extraneous load. The second explanation conjectured by Zacharia suggests that the haptic feedback needs to add significant value in order to be an effective facilitator.[4] This is in agreement with our finding that suggests that force could be derived visually from the spontaneous relaxation of the molecular structure. We suggest that the main reason why the haptic feedback hindered learning lies in the spontaneous relaxation of the molecular structure from which the force could be derived visually. The student could judge the extent of the repulsion based on how quickly they had to pull away in order to break the chemical bond without the rest of the molecule following. This appeared to be easier than interpreting the haptic feedback as indicated by the atomic movement theme. Potentially, the novelty of the haptic feedback drew the students' focus away from the visual representation and without being able to successfully interpret the force, this might have hindered learning, especially as we found close to no students who connected all three representations in one line of reasoning. This is aligned with the conceptual salience perspective that suggests that explicit attention to the information is needed in order for it to be processed.[226] In this case, the *conceptually salient* representation, meaning attention grabbing (in the case of the comparison group, but not in the experimental group) as well as carrying meaningful information, is the spontaneous movement of the spectator atoms as opposed to the potential haptic

feedback. In the study by Zohar and Levy, the atoms did not move spontaneously, which is one of the major differences between their and the present study design and therefore supports this explanation.[20] Similarly, in the study by Bivall and co-workers, the salient information was solely represented by haptic feedback, hence leading to the students who did not receive haptic feedback to not have access to this information.[180] The study by Murayama and co-workers differed both in the study design (I-(PS) vs. PS-I) as well as in the comparison condition (no intervention versus haptic device without haptic feedback).[111] We conjecture that as our comparison condition is stronger, the effect was reversed. We do not think that the study design is comparable in this case.

Although we could expect higher prior content knowledge from the university students investigated here compared to the high school students investigated in Zohar and Levy (2021),[20] the ambitious task of meaningfully connecting three representations of highly abstract concepts might need more scaffolding as we have seen over all themes a general trend towards heterogeneous conceptions and hence difficulties. However, even with optimized scaffolding, there still might be a remaining redundancy effect, as the force is represented both in the movement of the atoms and in the haptic feedback.[227] This effect suggests that redundant material might interfere with rather than facilitate learning as the coordination of multiple sources of information potentially increases cognitive load. Further research is necessary to explore the extent of this effect.

Generally, the guidance was kept intentionally low to get authentic feedback on the environment including the haptic feedback itself. We were interested in the way that the students would connect the multimodal representations and what conceptions they would activate during the experience, as well as in which would be challenged and which supported. Especially in the case of the haptic feedback, we found that more explicit scaffolding is needed, not only but also due to the way that we are talking about energy in colloquial language, both in German and English (as we discuss in the theme *energy is proportional to the force*). This is in line with the signaling effect as found for example in Mautone and Mayer.[228] They found that better transfer occurred when signals such as visual cues were included. In our case, no signaling was provided. Especially as prior conceptions of energy in a physics context might be activated, which includes energy as a stimulus,[229] the students might need a clearer distinction between the concept energy and force in this context. Otherwise, the conceptual mapping between the new information to the prior experiences (grounding[3]) will not be successful, prohibiting construction of coherent knowledge systems.[74]

The same argument is valid for the binary perception of chemical bonds. By better introducing how and why chemical bonds are displayed, the students' misconceptions

around chemical bonds could be better challenged and less of their resources would be needed to understand how and when chemical bonds are formed when we would like them to rather focus on the connection between energy and force.

Concerning the energy models activated by the students, we suggest that this rather informs the design of the instruction phase. As long as consolidation occurs, employing both models in the problem-solving phase should not create unnecessary difficulties. The instruction did not target the difference between energy levels and energy (hyper)surfaces and when it is appropriate to treat energy as a discrete and when as a continuous variable. This was not part of our intervention and further research is needed to understand how these conceptions are grounded and how they interact with each other to optimally improve instruction.

Finally, the framing of the study resulted in specific student mindsets. First, the students were told that they were piloting a learning environment that might become a part of the course for future generations. While this statement was an effective motivator for active participation, it potentially created suspicion of the learning environment and potentially made the students move from a learning practice towards a practice of critiquing and testing. Consequently, unexpected experiences could have two sources, either inaccurate preconceptions or a technical bug in the program. Second, for ethical reasons, the students were told that the experience will not affect their course grade and the learned content will not be explicitly examined. Potentially, this weakened their intent to learn and to make sense of the haptic feedback which then affected the learning outcome.[230] We increase the ecological validity in the follow-up study described in chapter 7 by better integrating the intervention into the course content.

5.4 Conclusion

Contrary to our hypothesis, we found a small hindering effect of haptic feedback on learning outcome. To explain this finding, we analyzed written answers to open-ended questions from the problem-solving phase. We concluded i) that the cognitive load might be too high in this case to connect the two visual representations and the haptic representation, ii) that multiple potentially incomplete conceptions around energy and its gradient were activated and that this should be considered carefully in the instruction phase, iii) that the haptic feedback would have to be introduced more explicitly, but that iv) even then, there might be a redundancy effect, as we found that the comparison group could interpret the force from the spontaneous movement of the atoms. We further discussed limitations of the study, with the major ones being the framing of the study as a test of the learning environment for future generations of chemistry students and

that the study content was not explicitly relevant for the grading of the course. We argued that this resulted in a student mindset that might have moved from a learning intention to testing or even gaming practice. In chapter 7, we investigate how scaffolding affects the learning outcome in the case of multimodal representations including haptic feedback. We conjecture that this further depends on the prior content knowledge including potential misconceptions in students.

6

Preparing for Learning Organic Chemistry with Haptic Feedback

We applied the identical design to the PFL study described in chapter 5 on second-semester students. The educational context was slightly different as the students had no quantum chemical target concept but were enrolled in an introductory organic chemistry course. The two-semester course was designed in a way that the students first built a toolbox of heuristics and models frequently employed by organic chemists for reaction elucidation. Only in the last quarter of the course, the students applied this toolbox to actual reaction mechanisms. At this point, between building the toolbox and being confronted with reaction mechanisms for the first time in their academic career, the study took place. In the following, the results of this study are presented. Moreover, in section 6.3, we contrast the results of the two PFL studies.

6.1 Method

The design and procedure were identical to the study with fourth-semester students described in chapter 5. However, the social context and the sampling differed. Specifically, the study was not set during a course, but the students were recruited on a voluntary basis and signed up to a session in their spare time. The students were grouped into sessions consisting of one to four students. In contrast, the fourth-semester students took part in a study during a laboratory course that they completed in groups of eight.

The study was approved by the ETH ethics committee and the study design, analysis plan, anonymized data as well as analysis scripts are available on OSF *.

6.1.1 Participants

$n = 24$ second-semester students (9 female, 15 male) participated in the study. The students were recruited via e-mail as well as in a short presentation during an introductory organic chemistry lecture. The students were enrolled in a bachelor program strongly related to chemistry (chemistry, biochemistry, chemical engineering or interdisciplinary sciences). Incoming characteristics per group are presented in Table 6.1.

6.1.2 Instruction phase

The instruction phase consisted of the lecture on mechanisms of substitution reactions. The problem-solving phase was tailored towards this instruction phase as the students explored two nucleophilic substitutions. However, the lecture was not changed in any way to take up the problem-solving phase material explicitly.

6.2 Results

Due to the recruitment strategy, the sample size was fairly small. Hence, the results lack statistical significance. In order to ensure completeness, the results are still presented in full.

The prior general chemical knowledge of the students was measured with the chemistry concept inventory by Mulford and Robinson.[231] The outcome of the latter was for no group significantly non-normal. Of all metacognitive variables, only the extraneous load in the experimental group was significantly non-normally distributed ($W = 0.91$, $p = 0.027$). No outcome or metacognitive variable showed significant difference in variance between groups. The assumption of homogeneity of the regression slopes was met as the interaction term between posttest performance and prior chemical knowledge was not statistically significant, $F(2, 18) = 0.811$, $p = 0.46$. Finally, an ANOVA was performed with the group as predictor for the covariate prior knowledge to ensure that the groups did not significantly differ in the covariate. There was no significant difference in prior chemical knowledge between groups, $F(2, 21) = 0.449$, $p > 0.1$, indicating that prior knowledge was a valid covariate.

Since all required assumptions were met, an ANCOVA was performed with the indepedent variable group, dependent variable posttest and covariate prior knowledge.

*https://osf.io/97yrd/?view_only=4f14de8516b84def9da5aaf04fac8014, accessed 2023-08-04

Table 6.1: Baseline characteristics for the participants. $N = 24$ ($n_{control} = 7$, $n_{comp} = 8$, $n_{exp} = 9$). Participants were on average 19.8 years old ($SD = 1.6$) and the groups did not differ in age. The control group completed the problem-solving tasks with a computer mouse, the comparison group with a haptic device with deactivated haptic feedback and the experimental group with a haptic device with activated haptic feedback. All characteristics are self-reported.

Baseline characteristic	Control		Comparison		Experimental	
	n	%	n	%	n	%
Gender						
Female	3	42.9	1	12.5	5	55.6
Male	4	57.1	7	87.5	4	44.4
Non-binary/ third gender	0	0.0	0	0.0	0	0.0
Prefer not to say	0	0.0	0	0.0	0	0.0
English proficiency						
Conversational	1	14.3	0	0.0	1	11.1
Fluent	6	85.7	6	62.5	6	66.7
Native Speaker	0	0.0	2	25	2	22.2
Gaming						
Less than once a year	4	57.1	1	12.5	4	44.4
Once a month	2	28.6	6	62.5	2	22.2
Once a week	1	14.3	1	12.5	3	33.3
More than once a week	0	0.0	0	0.0	0	0.0
Handedness						
Left	1	14.3	0	0.0	1	11.1
Right	6	85.7	8	100.0	8	88.9
Both	0	0.0	0	0.0	0	0.0

Neither prior knowledge ($F(1, 20) = 1.182$, $p = 0.290$, $\eta^2 = 0.064$) nor the condition ($F(2, 20) = 0.136$, $p = 0.873$, $\eta^2 = 0.013$) were significant predictors for learning outcome. This can be attributed to the small sample size. However, as prior knowledge is expected to have a large effect on learning outcome, we conjecture that general knowledge of chemistry might not be a necessary prerequisite for initial quantum chemistry learning.

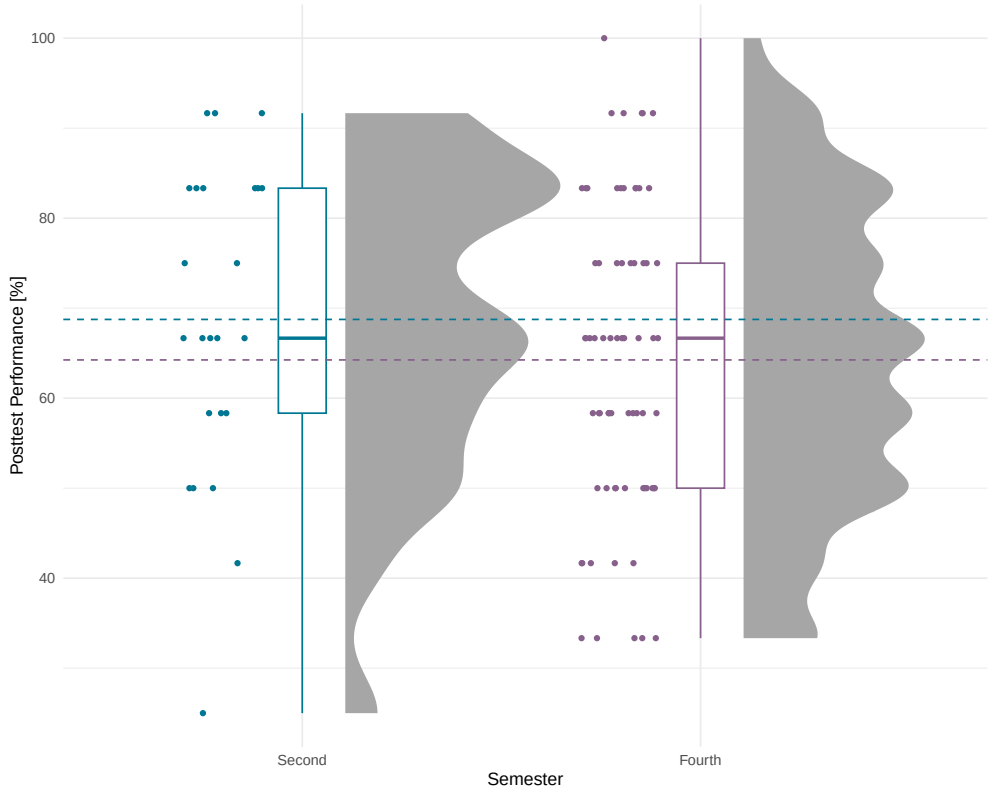


Figure 6.1: Distributions of posttest performance for the second- and fourth-semester participants. The means are indicated with dashed lines.

6.3 The impact of the recruitment strategy on study results

As the results of the second-semester students alone did not allow for any conclusions, these data were compared to the data collected in the study from chapter 5 in order to investigate the impact of the proficiency of the students in connection with the differing recruitment processes on the learning outcome.

When comparing the learning outcomes from the posttest, we found that the second- and fourth-semester students did not significantly differ in their posttest performance ($F(1.83) = 1.735$, $p = 0.191$, $d = -0.22$, see Figure 6.1).

Furthermore, the analysis of the performance on the individual posttest items with Fisher's exact test revealed that only in two out of the twelve items, the students of the second-semester chose significantly different answer options than their colleagues in the fourth-semester. When being asked, "If a random bond in a stable molecule is being broken, how will the potential energy of the system change?", the second-semester students mostly chose the correct answer "It will increase", while a significant amount of fourth-semester students chose the option "If it decreases or increases will depend

on how much energy is stored in the bond" ($p = 0.008$). Second, when being asked to indicate what points out of a selection on a potential energy surface corresponded to a transition state structure, second-semester students more often chose the incorrect option of random points on a surface not belonging to a minimum, maximum or saddle point, while the fourth-semester students mostly chose the correct saddle points ($p = 0.045$). Refer to the Appendix A for the exact wording of the test items.

While these differences might indicate persisting and even developing misconceptions about the concept of the chemical bond (In chapter 8, an in-depth qualitative analysis on chemical bond conceptions is presented) or difficulties to integrate basic quantum chemical models into their prior conceptions, there are many limiting factors to this analysis. First, the second-semester students were recruited on a voluntary basis while the fourth-semester students were enrolled in a course during which the study took place and only had to indicate whether they gave consent for the data to be used for research. Therefore, we could expect mainly highly motivated and ambitious second-semester students with a tendency to be interested in computational chemistry to participate while we worked with a more heterogeneous sample of fourth-semester students. Second, the foci of the studies differed. While the fourth-semester students were being prepared to conduct their first computational chemistry calculations, the second-semester students focused on the nucleophilic substitution reaction presented to them. Therefore, prior knowledge relevant to the posttest might have been more or less activated, depending on the study.

Table 6.2: Fit evaluation for models with differing numbers of clusters. LL denotes the maximum Logarithmic Likelihood,[232] BIC the Bayes Information Criterion,[233] AIC the Akaike Information Criterion[234] and df the residual degrees of freedom. The two-cluster model displays the **lowest value** for AIC, indicating that the data are best described by two answering clusters besides the non-informative single cluster.

	LL	BIC(LL)	AIC(LL)	df
1-Cluster	**-934.23**	**2026.72**	1938.46	57
2-Cluster	-881.26	2083.57	**1904.52**	21
3-Cluster	-851.24	2186.31	1916.48	-15
4-Cluster	-834.07	2314.75	1954.14	-51

Finally, as the semester did not significantly predict performance, we investigated if there were any other distinct answering patterns and whether they belonged to certain groups of students. To that end, we performed a latent class analysis[235,236] (LCA). As we did not theoretically expect a specific number of clusters, we conducted LCAs of one

to four clusters (see Table 6.2). We found that answering patterns could best be divided into two groups which corresponded to higher and lower achievers. Remarkably, these groups had similar numbers of fourth and second-semester students as well as equal parts of the three study conditions (see Figure 6.2).

Again, we conjecture that this finding is related to the differing recruitment strategies and the consequently biased sample for the study targeting second-semester students. This stresses the importance of authentic class room studies that do not rely on voluntary participation but include the study in the regular class room activities.

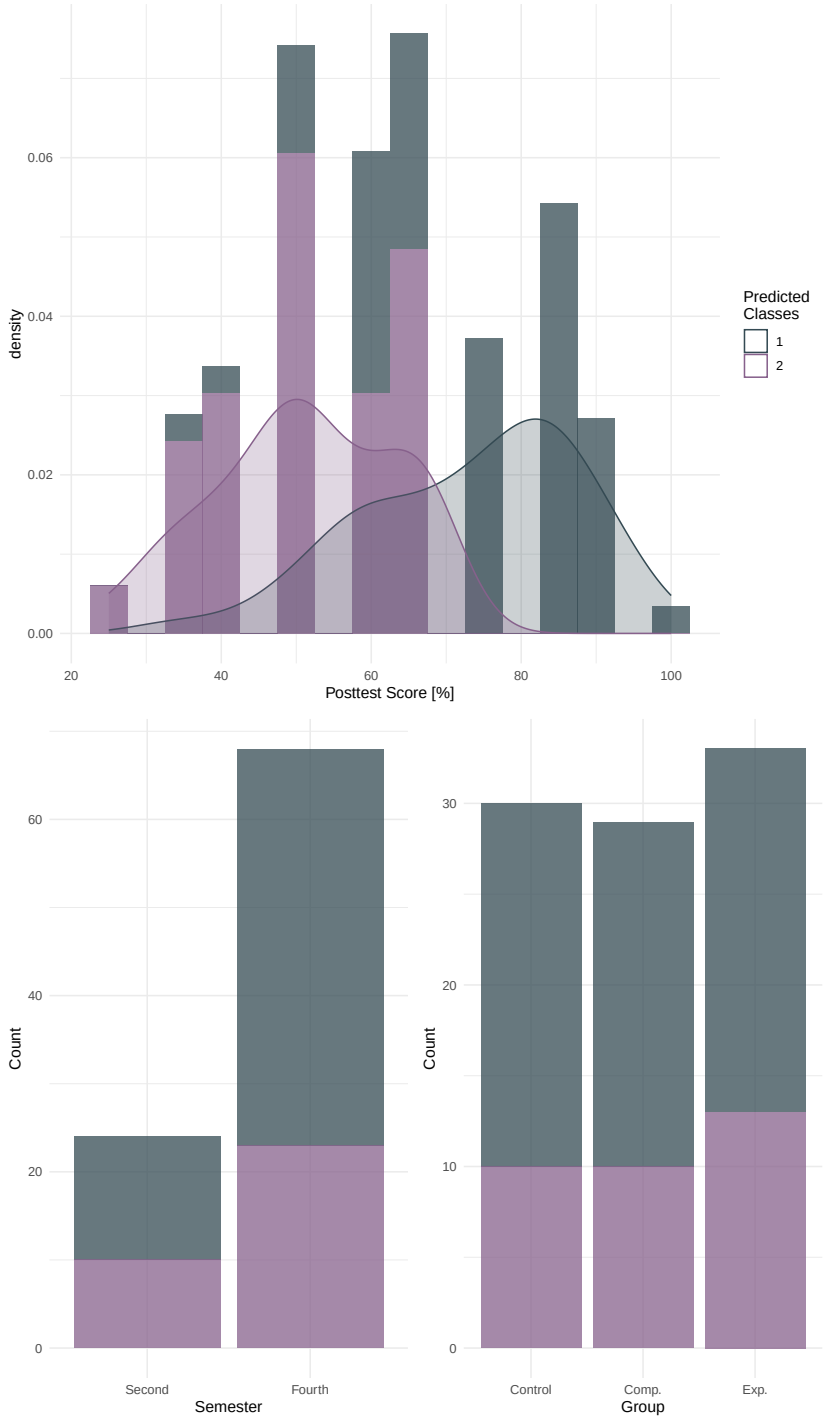


Figure 6.2: Latent class analysis resulted in two groups of students who showed distinct answering patterns. These groups did not correspond to the semesters or to the conditions of the studies, but to higher and lower achieving students.

7

The Effect of Explicit Embodied Metaphors on Learning Outcome

In this chapter, we present a follow-up study to the study discussed in chapter 5 which resulted in lower performance for students receiving haptic feedback compared to students who did not. We conjectured that one main limitation of the study design in chapter 5 was the strong separation between course content and study material. While we aimed to build intuition relevant to the course content, the students might not have grasped this connection and might have participated in the study without intent to learn and instead with the intent to improve the environment. Second, we argued that the haptic feedback was not beneficial to the students since they were not able to successfully map the sensory experience to the new domain of quantum chemistry. Here, we intended to improve on these two points by better incorporating the study into the course content (see subsection 7.1.2) and by making the target metaphor explicit to the students. The precise hypotheses are given in section 2.3.

7.1 Methods

7.1.1 Participants

The study was conducted one year after the one presented in chapter 5 in the context of the same course. Hence, again, the participants consisted of chemistry, chemical engineering and interdisciplinary science fourth-semester undergraduate students ($n = 48$,

Table 7.1: Baseline characteristics for the participants. $N = 48$ ($n_{comp} = 24$, $n_{exp} = 24$). Participants were on average 20.9 years old ($SD = 1.4$) and the groups did not differ in age. The comparison group received a visual representation of the force, the experimental group a haptic representation. All characteristics are self-reported.

Baseline characteristic	Comparison		Experimental	
	n	%	n	%
Gender				
Female	7	29.2	10	41.7
Male	16	66.7	14	58.3
Non-binary / third gender	0	0.0	0	0.0
Prefer not to say	1	4.2	0	0.0
English proficiency				
Conversational	9	37.5	5	20.8
Fluent	12	50.0	14	58.3
Native Speaker	3	12.5	5	20.8
Gaming				
Less than once a year	10	41.7	10	41.7
Once a month	6	25.0	7	29.2
Once a week	4	16.7	4	16.7
More than once a week	4	16.7	3	12.5
Handedness				
Left	1	4.2	5	20.8
Right	23	95.8	19	79.2
Both	0	0.0	0	0.0

35% female, 63% male, 0% non-binary/third gender, 2% preferred not to say). The participation in the course was mandatory, but data were solely collected if informed consent was given. Consent was sought and received from all. Participants were randomly assigned to one of two groups depending on which week of the semester they signed up to the course. Some of the students have learned about reaction mechanisms generally and the Diels-Alder reaction mechanism specifically approximately one year prior to the course. However, it is important to note that not all students have taken the general organic chemistry exam. We further assumed that they were familiar with the static two-dimensional energy graph that depicts reactions as traditionally found in text books. The students did not yet explicitly learn how force is related to the energy and they were not yet aware of the inherent multidimensionality of the potential energy

hypersurface according to the curricula of their programs. Moreover, as mentioned in previous chapters, they potentially activated prior conceptions of energy in a physics context (see subsection 3.2.2 for an overview of the empirical literature on relevant student conceptions). Table 7.1 gives an overview of the baseline statistics. The study was approved by the ETH ethics committee and the study design, analysis plan, anonymized data as well as analysis scripts are available on OSF *.

7.1.2 Material

Set-up

The set-up was highly similar to the previous study described in chapter 5. However, few adjustments were made based on the previous results. First, a general limitation of the previous study was the explicit phrasing of it as a study and the clear distinction between the course content and the study content. Here, we adapted the learning material to be better embedded in the general course content. Specifically, we changed the example reaction to the reaction that they would subsequently investigate using traditional electronic structure calculations employing Orca.[237,238] Specifically, they conducted a Diels-Alder reaction interactively and saved the final product to later further optimize it with density functional theory (DFT). An excerpt form the instructions is given in Figure 7.1. For better inclusion of the study content into the course content, the study instructions were included in the course instructions for this iteration of the study.

*https://osf.io/c23rx/?view_only=54c73f4b83fa48ac87b35f40f9ff879d, accessed 2023-08-04

und versuchen Sie die Reaktion mit dem haptischen Gerät auszuführen, indem Sie die Auswahl langsam auf den anderen Reaktanden zubewegen.

Eine höhere Abstossung steht in Verbindung zu einem steileren Anstieg der Energie (und analog eine höhere Anziehung mit einem steileren Abstieg), so etwa, wie wenn Sie einen steileren Berg erklimmen. Der Aufwand, um den Gipfel zu erreichen, ist höher, als bei einem Hügel. Sie fühlen die Kraft, welche auf das gewählte Atom wirkt.

7. Sobald Sie das gewünschte Produkt erzeugt haben, notieren Sie sich die Reaktionsenergie. Notieren Sie sich ausserdem qualitativ, wie einfach es war, die Reaktion durchzuführen. Was ist Ihnen aufgefallen? Sie werden später diese Notizen in der Gruppe vergleichen. Sie könne ausserdem durch anwählen des entsprechenden Knopfes oben rechts zurückspulen und an verschiedenen Stellen die Orbitale anzeigen lassen, oder die Repräsentation der Atome ändern (z.B. zu atomaren Ladungen). Machen Sie sich auch dazu Notizen. Diese werden am Schluss zu Forschungszwecken abphotograhiert. Falls Sie dazu Fragen haben, sprechen Sie uns gerne darauf an.

und versuchen Sie die Reaktion mit dem haptischen Gerät auszuführen, indem Sie die Auswahl langsam auf den anderen Reaktanden zubewegen.

Die Pfeile geben Ihnen die Kräfte an, welche auf die individuellen Atome wirken. Längere Pfeile stehen in Verbindung zu einem steileren Anstieg oder Abstieg der Energie.

7. Sobald Sie das gewünschte Produkt erzeugt haben, notieren Sie sich die Reaktionsenergie. Notieren Sie sich ausserdem qualitativ, wie einfach es war, die Reaktion durchzuführen. Was ist Ihnen aufgefallen? Sie werden später diese Notizen in der Gruppe vergleichen. Sie könne ausserdem durch anwählen des entsprechenden Knopfes oben rechts zurückspulen und an verschiedenen Stellen die Orbitale anzeigen lassen, oder die Repräsentation der Atome ändern (z.B. zu atomaren Ladungen). Machen Sie sich auch dazu Notizen. Diese werden am Schluss zu Forschungszwecken abphotograhiert. Falls Sie dazu Fragen haben, sprechen Sie uns gerne darauf an.

8. Speichern Sie die Struktur mit einem sinnvollen Namen ab.

Figure 7.1: Excerpt from the instructions for the preparatory intervention, left for the experimental group receiving haptic feedback and right for the comparison group receiving visual feedback. In bold, the explicit explanation of the representation is given. For the experimental group, this translates to "A bigger repulsion corresponds to a steeper ascend in energy (and similarly a bigger attraction to a steeper descend), similarly to when you are climbing a steeper mountain. The effort needed to reach the peak is higher than if it was a hill. You feel the force, which acts on the selected atom". For the comparison group, this translates to "The arrows correspond to the forces, which act on the individual atoms. The longer the arrows, the steeper the ascend or descend of the energy". The students conducted a Diels-Alder reaction for which they later recalculated the relevant electronic structures.

The change from nucleophilic substitution reactions to the Diels-Alder reaction had two consequences. First, the students only had to perform one reaction with a clear goal to further use the product in later calculations. Second, this task was harder since the Diels-Alder reaction coordinate consisted of two carbon atoms simultaneously moving towards two other carbon atoms. This simultaneous movement was possible by selecting multiple atoms, but it was quite difficult and the students needed multiple attempts to succeed. Together, this led to the two studies taking a similar amount of time.

There was no instruction phase in the form of a video such as in chapter 5. However, we expected the students to prepare during the problem-solving phase for the traditional computational studies that they conducted later on. Traditionally, the electronic structure calculations work as a black box for novices in computational chemistry and the aim of this intervention was to provide the students with a connection to the physical phenomenon behind this black box.

Finally, a newer version of the SCINE Heron interface was used here; however, not much changed for the students except for the increased robustness of the software.

Measures

We investigated the effect of receiving a grounding opportunity in embodied experiences relative to receiving visual information of the same concept on learning outcome, learning mechanisms and motivational variables. Therefore, the dependent variable was performance which was measured through an in-house developed posttest. The multiple-choice posttest was validated by two chemists and was refined three times, once by online distribution to students, once based on the results in a previous usability study for SCINE Heron (see chapter 4) and once in the prior PS-I studies (see chapters 5 and 6). The test measured multiple interconnected concepts such as a change of potential energy related to molecular structure, change of energy related to force, energy difference related to reactivity or energy extrema as critical points during reactions (see Appendix A for detailed learning objectives and test items).

Furthermore, cognitive, affective and motivational variables were included according to the stated hypotheses (see section 2.3). An overview of the measures and reliability is given in Table 7.2.

Table 7.2: Overview of learning mechanism and content knowledge measures. The response rate of the posttest was too low to calculate the total omega.

Construct	Example item	Type	Total Omega	Source
Extraneous load	It was easy for me to distinguish important from unimportant information.	Likert scale	0.93	Original items from Glogger-Frey et al.[220]
Positive affect	Interested	Likert scale	0.92	Original items from Watson et al.[218]
Knowledge gap awareness	My knowledge was insufficient to carry out the tasks.	Likert scale	0.91	Original items from Gloger-Frey et al.[220]
State curiosity	I want to know more.	Likert scale	0.91	Original items from Naylor.[219]
Attention	The style of the writing is boring.	Likert scale	0.91	Original items from Keller.[112]
Relevance	It is clear to me how the content of this material is related to things I already know.	Likert scale	0.90	Original items from Keller.[112]
Confidence	When I first looked at this lesson, I had the impression that it would be quite easy for me.	Likert scale	0.89	Original items from Keller.[112]
Satisfaction	It felt good to successfully complete this lesson.	Likert scale	0.93	Original items from Keller.[112]
Content knowledge	What do individual points on a potential energy graph correspond to? a. Different spatial arrangements of atoms. b. Different levels of excitation. c. Different vibrational frequencies. d. Different atomic velocities.	Single-choice	-	Developed in-house.

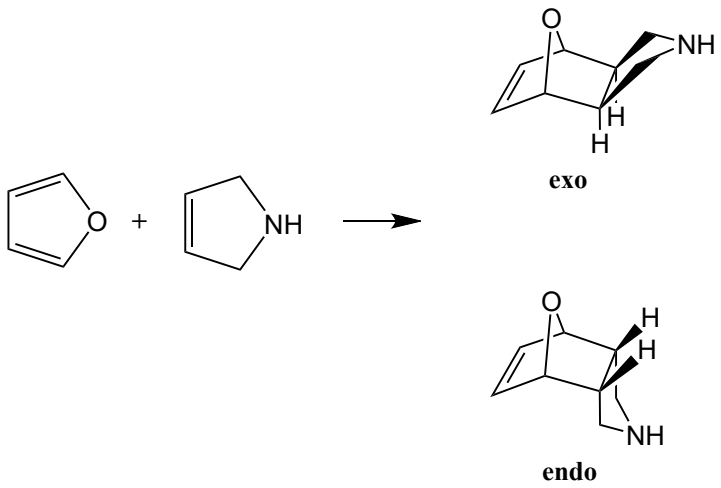


Figure 7.2: The Diels-Alder reaction of furan and 3-pyrroline that the students were asked to perform interactively. This figure is reproduced from the course instructions.

7.1.3 Design

We followed a between-subjects preparatory problem-solving design with two groups differing in the problem-solving mode. The experimental group received haptic feedback during problem-solving as well as a written explanation of the target conceptual metaphor. The comparison group performed the same tasks but received visual indication of the force as arrows on the atoms instead of haptic feedback. They further received written explanation of the representation, but here, this explanation was of literal instead of metaphorical nature (see Figure 7.1 for the exact wording of the instructions).

7.1.4 Procedure

Again, the experimental study followed a preparation for future learning (PFL) design.[9,16] Prior to the course, the students filled out a demographic questionnaire. In the course, they started by first building the relevant structures with Avogadro.[239] As in the study described in chapter 5, they then familiarized themselves with the environment with simple navigation tasks such as rotating, zooming or dragging and dropping of an atom around another molecule. After self-proclaimed proficiency, they moved on to the problem-solving phase in which they were asked to perform the Diels-Alder reaction depicted in Figure 7.2.

After successfully creating the product, they were asked to fill out the questionnaire measuring the motivational and affective variables. They then continued the course by optimizing the relevant structures of the Diels-Alder reaction with DFT. The procedure

Pre-Measurements	Problem-Solving		Post-Measurement I	Instruction	Post-Measurement II
Demographic data Exam grades	No HF	HF	Affective & cognitive questionnaires	Electronic structure calculations	Posttest
	Comparison	Experimental			

Figure 7.3: PFL design. HF denotes the haptic feedback. Post-measurement I occurred immediately after the PS phase and post-measurement II was administered towards the end of the semester to ensure similar proficiency in all students.

took approximately 90 minutes out of the total 8 hours duration of the course. An overview is given in Figure 7.3.

The posttest that measured their understanding of the target concepts (reaction energy, activation energy and reaction coordinate) was sent out to all via e-mail towards the end of the semester.

7.1.5 Analysis

Quantitative analysis

Hypothesis 1 was tested in an ANCOVA with dependent variable posttest, independent variable group and covariate prior content knowledge.

Hypothesis 2 was tested in a parallel mediator analysis with the PROCESS macro in R 2022.12.0.[222] We included learning mechanisms that are known for being salient in PS-I, knowledge gap awareness, positive affect and state curiosity as mediators in the model as well as prior content knowledge as covariate.

Finally, the ARCS variables[112] were analyzed with the robust Yuen's test for trimmed means as we expected the factors to be independent.[240]

7.2 Results

The physical chemistry grade of the first year introductory course was taken as proxy for relevant prior content knowledge as several students did not participate in the other general chemistry lectures (inorganic and/or organic chemistry). The prior knowledge was non-normally distributed in the comparison group ($W = 0.91$, $p < 0.05$). Furthermore, confidence was non-normally distributed in the comparison group ($W = 0.82$, $p < 0.001$). However, confidence and prior knowledge were not significantly correlated ($r(45) = 0.33$, $p > 0.5$). All other variables were normally distributed. Hence, we intended to employ robust methods to perform the quantitative analyses. The assump-

tion of homogeneity of variance (except for relevance, $F(1, 46) = 4.4$, $p < 0.05$) and the assumption of homogeneity of regression slopes ($F(1, 8) = 0.23$, $p = 0.64$) were met. The assumption of independence of prior content knowledge and treatment effect was not met as the groups significantly differed in their prior knowledge ($F(1, 45) = 9.53$, $p < 0.005$). Due to the unmet requirements, we intended to conduct a robust ANCOVA. Finally, the point in time when the students participated in the study and hence the time difference between the intervention and the posttest was not a significant predictor for posttest performance ($F(1, 11) = 0.36$, $p = 0.56$). Descriptive statistics are given in Table 7.3.

Table 7.3: Overview of descriptive statistics of the continuous variables. Prior knowledge is an incoming characteristic, the others are outcome variables. All values are given in percent. $\tilde{\mathbf{x}}$ denotes the median, $\bar{\mathbf{x}}$ the mean and **SD** the standard deviation.

Variable	Group	n	Min	Max	$\tilde{\mathbf{x}}$	$\bar{\mathbf{x}}$	SD
Prior knowledge	Exp.	24	54.2	100.0	75.0	77.8	10.7
	Comp.	23	58.3	87.5	66.7	69.2	8.1
	all	47	54.2	100.0	75.0	73.6	10.4
Extraneous load	Exp.	24	50.0	91.7	68.8	70.0	9.9
	Comp.	24	50.0	87.5	62.5	66.0	11.4
	all	48	50.0	91.7	66.7	68.0	10.7
State curiosity	Exp.	24	58.8	87.5	74.4	74.4	8.4
	Comp.	24	47.5	82.5	62.5	63.6	9.9
	all	48	47.5	87.5	68.8	69.0	10.6
Knowledge gap awareness	Exp.	24	20.0	66.7	43.3	43.3	12.4
	Comp.	24	26.7	76.7	43.3	46.1	14.4
	all	48	20.0	76.7	43.3	44.7	13.4
Positive affect	Exp.	24	46.0	84.0	69.0	67.2	11.9
	Comp.	24	28.0	74.0	56.0	54.6	12.9
	all	48	28.0	84.0	59.0	60.9	13.8
Attention	Exp.	24	53.3	91.7	80.8	78.0	11.4
	Comp.	24	50.0	90.0	71.7	71.4	10.6
	all	48	50.0	91.7	75.8	74.7	11.4
Relevance	Exp.	24	46.7	95.6	74.4	74.0	14.2
	Comp.	24	51.1	86.7	67.8	67.4	9.5
	all	48	46.7	95.6	70.0	70.7	12.4
Confidence	Exp.	24	37.8	84.4	68.9	66.3	13.0
	Comp.	24	0.0	82.2	65.6	62.7	17.4
	all	48	0.0	84.4	66.7	64.5	15.3
Satisfaction	Exp.	24	31.1	62.2	46.7	48.3	9.9
	Comp.	24	24.4	62.2	44.4	43.9	9.6
	all	48	24.4	62.2	45.6	46.1	9.9
Posttest	Exp.	7	41.7	83.3	58.3	61.9	14.3
	Comp.	5	58.3	91.7	83.3	76.7	13.7
	all	12	41.7	91.7	66.7	68.1	15.4

7.2.1 Performance

As expected, the response rate for the posttest that was administrated only towards the end of the semester via e-mail was quite low ($n = 13$). Consequently, the results did not reach statistical significance and the sample size was not sufficient to perform the robust ANCOVA at all, as this method requires multiple observations at several values of the covariate, here the posttest performance. The standard ANCOVA did not show significant differences between groups ($F(1, 9) = 2.31$, $p = 0.16$, $d = -1.05$).

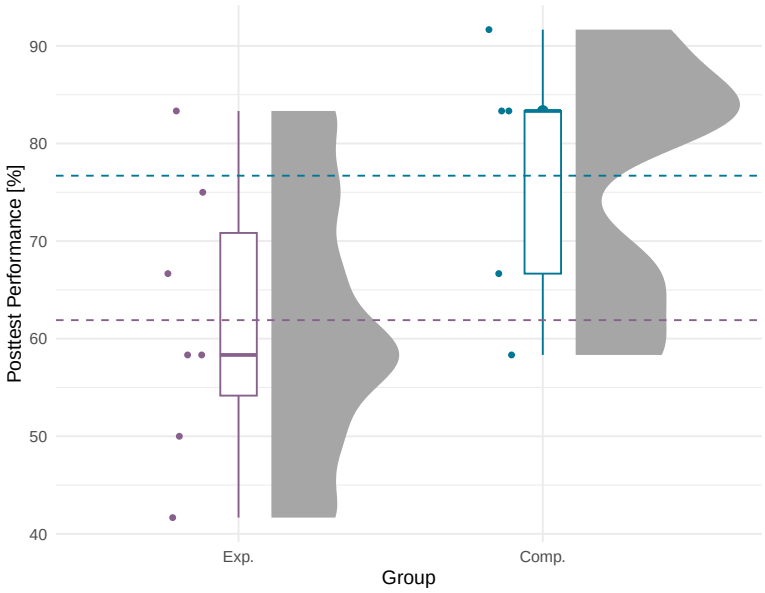


Figure 7.4: Distributions of posttest performance for the individual groups. The means are indicated with dashed lines.

For the same reason, a mediator and moderator analysis was also not feasible. While there are significant differences in some of the learning variables expected in PFL designs, specifically positive affect and state curiosity (see Table 7.4), it could not be concluded whether these mechanisms were effective mediators of the learning process.

7.2.2 Motivational factors

An overview of all comparisons of affective, cognitive or motivational variables is given in Table 7.4 and graphically in Figure 7.5. From the ARCS model,[112] we found significant differences between groups in the perceived relevance ($t(23.13) = 2.08$, $p < 0.05$) and attention ($t(26.33) = 2.07$, $p < 0.05$).

Table 7.4: Overview results of robust t-tests as well as robust effect size measures ξ[225]. Results with p-values < 0.05 are indicated in **bold**.

Construct	Test statistic	df	ξ	ξ 95% confidence interval	p-value
Extraneous load	1.37	28.59	0.30	[0.00, 0.65]	0.181
Positive affect	3.28	29.29	0.55	[0.21, 0.83]	0.00266
Knowledge gap awareness	0.458	28.11	0.093	[0.00, 0.46]	0.650
State curiosity	3.70	29.95	0.64	[0.38, 0.92]	0.00086
Attention	2.07	26.33	0.41	[0.023, 0.79]	0.0482
Relevance	2.08	23.13	0.43	[0.051, 0.74]	0.0489
Confidence	0.822	28.01	0.15	[0.00, 0.55]	0.418
Satisfaction	1.26	26.99	0.25	[0.00, 0.60]	0.217
Content knowledge	1.71	4.36	0.63	[0.00, 0.92]	0.157

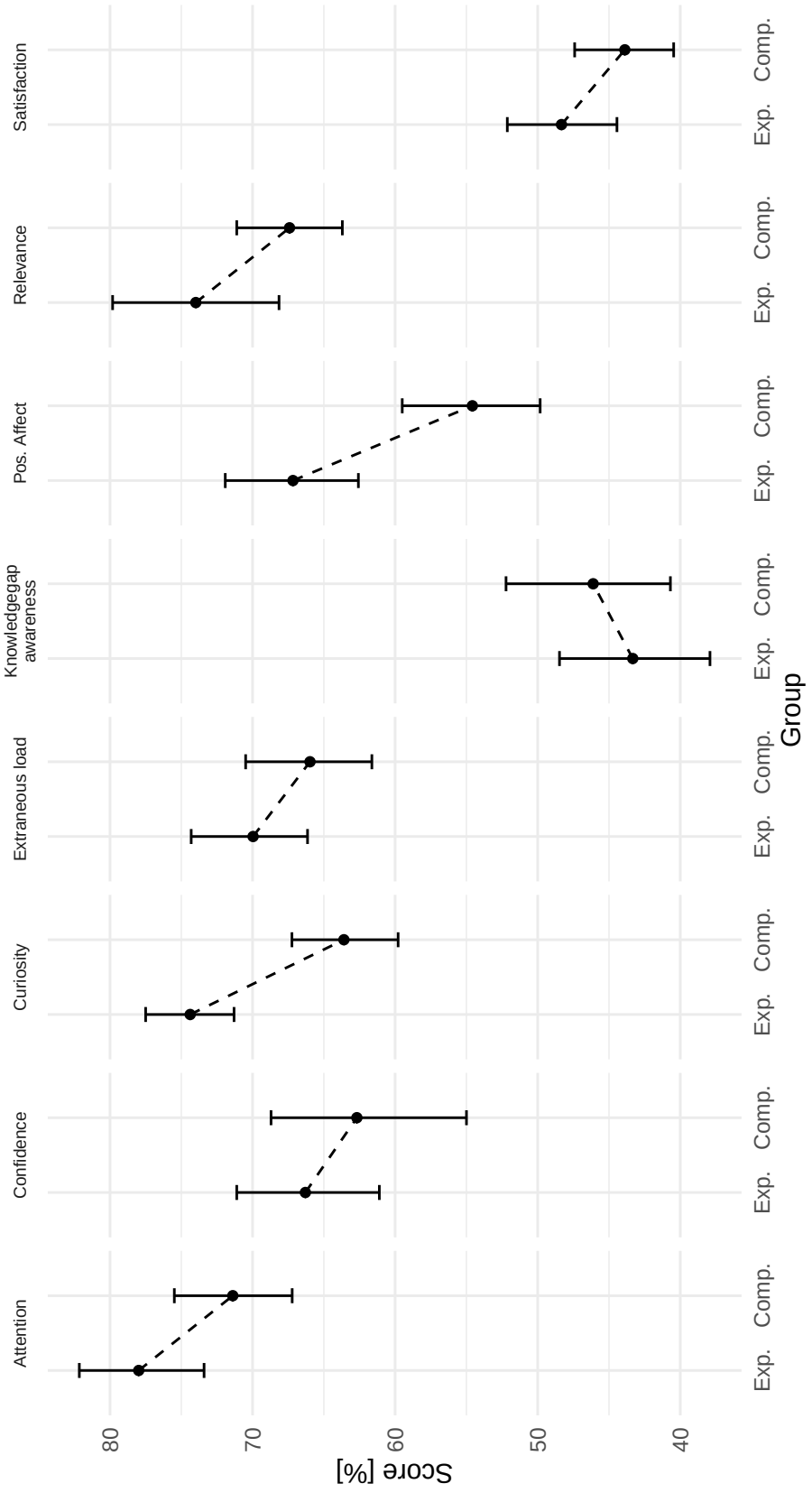


Figure 7.5: Group differences in the measured learning mechanisms. This includes cognitive and affective mechanisms frequently associated with preparation for future learning designs as well as motivational factors.

7.3 Discussion

In this chapter, we compared receiving haptic feedback together with explicit explanation of the conceptual metaphor to receiving a visualization that was less embodied. We found increased attention, curiosity and perceived relevance in the students who received haptic feedback. As the response rate to the posttest was quite low, we could not draw any inferences about the performance difference between the groups. Moreover, the responses that we did receive can be attributed to a specific subgroup of students who were motivated to participate, eager to learn and generally interested in the learning content.

Concerning the learning mechanisms associated with PS-I, knowledge gap awareness, positive affect and state curiosity, we found that the experimental group showed higher state curiosity immediately after the intervention. Moreover, we found that positive affect was enhanced in students who received metaphorical explanation as opposed to the technical explanation that the comparison group received (see Figure 7.1 for the exact wording of the instructions). The metaphor was made explicit by comparison to an experience that the students potentially could relate to, a hike to the peak of a mountain. This authentic storyline meets the design criterion *affective draw*[19] for productive failure,[16,18,105] a special case of PS-I, which is associated with higher curiosity and positive affect.

We further considered motivational factors, specifically attention, perceived relevance, confidence and satisfaction. We found that the experimental group perceived the intervention to be more relevant to their personal situation. This finding is aligned with previous results by Chatain and co-workers.[82] In their study targeting the max-flow problem of graph theory, they compared students in an abstract, pen-and-paper condition to students who manipulated the graph on a tablet and to students who interacted with a graph in an embodied fashion in virtual reality. They denoted the latter two as manipulated and embodied concreteness, respectively. These two aspects of concreteness evoke different cognitive learning mechanisms. Embodied representations may activate direct state induction, modal priming and sensorimotor simulation while manipulation-based representations may activate sensorimotor simulation only.[82,85] For our study, mainly the mechanisms modal priming and sensorimotor simulation were of importance as we were not aiming to activate any specific feelings (direct state induction) in the participants through bodily actions. Modal priming describes one direction of the bidirectional mechanism cognitive-sensorimotor transduction described in the grounded and embodied learning framework (see section 2.2).[3,50] That is, the embodied interaction can activate abstract concepts, such as in our case the quantum chemical force or in the case of Chatain and co-workers the max-flow concept of graph theory. Chatain and co-

workers argued that precisely due to this modal priming that is exclusively activated in embodied interaction, the perceived relevance of the target concept is increased. They argued that the embodied experience allowed the students to reconnect the learning content – which is not experiential for both their and our study – to prior (bodily) experiences, which is a crucial part of relevance in a learning context.[241] Rephrasing this in the framework applied throughout this work, the students successfully mapped their prior experience onto the quantum chemical domain and hence successfully constructed a conceptual metaphor.

Finally, we found increased attention in the experimental group. As attention integrates arousal theory, boredom, sensation seeking and, most importantly for this finding, curiosity, this was to be expected.[112] Furthermore, we conjecture that as the construction of the conceptual metaphor was more successful than in previous studies, the offloading mechanism (see chapter 2.2) was activated and with the lowered cognitive load, attention might have been increased.

While we were not able to show the hypothesized significant difference in performance between the groups, we were able to successfully implement the affective draw criterion for productive failure designs indicated by a significant difference in curiosity and positive affect.[19] Moreover, we operationalized the embodied and manipulated concreteness described by Chatain and co-workers[82] in a quantum chemical context instead of a math context and with haptic interaction instead of movement in immersive virtual reality and found that similarly, perceived relevance increased for the experimental group.

The main limitation of this study was the time constraint. The students worked with the haptic device for 60 to 120 minutes. For some of them, the new tool in addition to the new, unfamiliar and highly abstract quantum chemical concepts, was too confusing. Furthermore, while we decreased their awareness that this was a study and not relevant to the course grading, they were still aware that this was part of a doctoral project and therefore some of them still expressed distrust in the environment when it did not respond as expected, for example if the number of bonds on a certain atom did not correspond to the octet rule. Finally, the students participated in the study as groups of eight and while this was anecdotally less of an issue in the prior study described in chapter 5, we found that in this case, the group dynamics were highly heterogeneous ranging from highly focused and motivated to chaotic and negative. However, these factors have not been measured and are solely anecdotal evidence.

7.4 Conclusion

Due to the low response rate to the posttest, we are not able to draw inferences about the performance of the students and hence cannot comment on the hypothesis that successfully grounding energy change in prior bodily experience leads to better performance in a posttest than having a visual indication of energy change that is less grounded. However, we found that some of the learning mechanisms that were expected to have an impact on performance differed significantly between the two groups. Specifically, the state curiosity and positive affect as well as perceived relevance and attention were higher for the experimental group immediately after the intervention. This implies that explicit metaphors might comply with the design principle *affective draw*[19] and that therefore, further exploration of the role of affective variables in the learning process involving explicit conceptual metaphors might be of interest. Furthermore, we found that modal priming was a crucial mechanism to increase perceived relevance of the learning content.[82] While the results of the learning mechanisms showed a general trend in favour of the conceptual metaphor condition, the statistically insignificant result of the posttest showed a large effect in favour of the literal, visual representation. Although this result is not reliable in itself, future research should not only reproduce this study with higher statistical power, but also question the conceptual metaphor used. Specifically, climbing a mountain is more exhausting than climbing hill – not necessarily due to the steeper slope but potentially also due to the higher peak. If the students focused on the peak, this might support the misconception that the force is proportional to the energy (as opposed to proportional to the derivative of the energy).

8

The Multiple Student Conceptions of the Chemical Bond

The chemical bond is described with various models ranging from two electrons being shared between two nuclei to the energetically minimum distance between two atoms. How we visualize these bonds further primes students to select one model over another. Previous research has targeted misconceptions of the chemical bond, but how these conceptions interact and evolve remains relatively underexplored. In this chapter, we examine the diversity of conceptions in undergraduate chemistry students who have started to learn about quantum chemistry. We present results of a thematic analysis of data produced by students between the first and fourth-semester while interacting with a chemical simulation. We find that the chemical bond was largely conceptualized either as a rigid entity or in relation to energy. This is in agreement with the most commonly employed models in the scientific community. However, we argue that the students struggle to understand the contextual dependency and the plurality of equally valid models.

8.1 Methods

We triangulate qualitative data from three studies with university students, that is think-aloud data from a user study and written answers to open questions from two preparation for future learning (PFL[16,18,104,105]) studies. In the following, the relevant aspects of the three studies are described and the analysis process is specified. Study I

denotes the user study with first and third semester students (see chapter 4) and study II (see chapter 6) and study III (see chapter 5) denote the PFL studies with second and fourth-semester students respectively. For a more detailed account of the studies, refer to the respective chapters.

8.1.1 Participants

All three studies were conducted with undergraduate students from ETH Zurich who were enrolled in a program strongly related to chemistry (biochemistry, chemistry, chemical engineering or interdisciplinary sciences). An overview of the measured traits is given in Table 8.1. The proficiency ranged from being a first semester student in a general chemistry lecture without any prior knowledge in quantum chemistry to fourth-semester students attending their first computational chemistry course. The participants for study I and II were recruited on a voluntary basis via e-mail. For study II, we additionally advertised the study in a general chemistry lecture. Participants of study III participated in the study as part of a course. For all studies, the data were only used if the student gave informed consent. All studies were approved by the ethics committee of ETH as proposal EK-2021-N-41.

Table 8.1: Overview of income characteristics of participants as well as obtained data for the three studies. The descriptive count data relates to the students whose utterances were coded at least once with a code relevant to the chemical bond concept.

Characteristic	Study I	Study II	Study III
Semester	1 & 3	2	4
N_{total}	7 & 3	24	68
N_{used}	4 & 1	11	15
English fluency (conversational, fluent, native)	0, 3, 2	1, 8, 2	2, 10, 3
Gender (f, m, x, -)	1, 4, 0, 0	4, 7, 0, 0	7, 7, 1, 0
Age	20.6 ± 3.0	19.7 ± 1.4	20.9 ± 1.2
Data	Think-aloud	Written answers to open-ended questions	Written answers to open-ended questions

8.1.2 Materials

All studies employed the same learning environment, SCINE Heron. SCINE Heron is described in more detail in chapter 4. Importantly for the presented analysis, the

molecules are represented by sticks and the atoms by balls. Chemical bonds are displayed based on the distance between atoms. This choice of visualization results in i) the breaking and forming point of the bond being somewhat arbitrary from a chemical point of view and ii) arbitrarily many bonds per atom as opposed to as much as the octet rule would allow for molecules made from second-period atoms.

The participants of study I interacted with a pilot version of the final interface while the participants of study II and III both interacted with the final product. While the visuals differed in some aspects as can be seen in Figure 8.1, the visualization of the bond did not change.

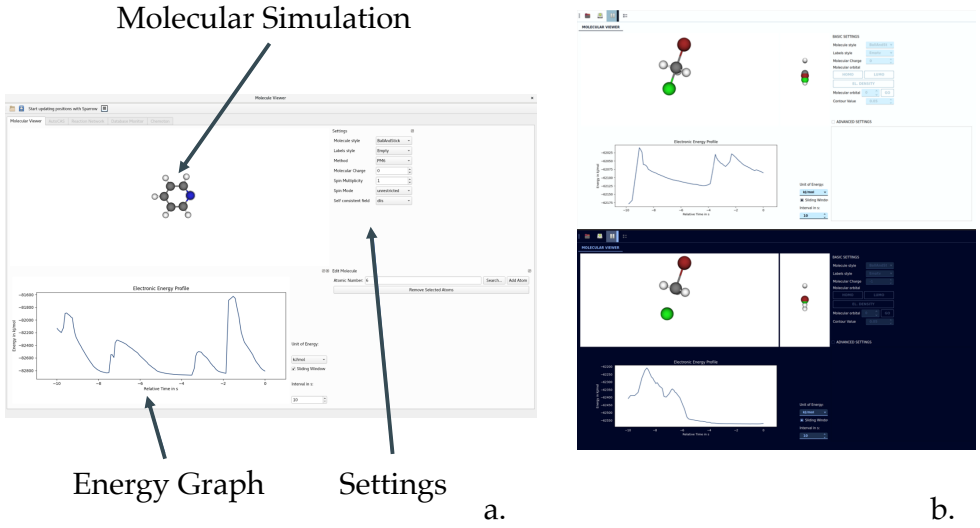


Figure 8.1: SCINE Heron[49] at the stage of the usability study (left) and at the preparation for future learning studies (right). In both versions, the chemical bonds are displayed based on distances between atoms.

In all studies, the tasks were highly similar. First, the students were familiarized with the environment by performing navigation tasks such as rotating, zooming, dragging and dropping. After self-proclaimed proficiency, the students were presented with three tasks. First the students broke the bonds of chloromethane. In this task, the students were prompted to compare bond strengths and energy change. The second and third task consisted of two nucleophilic substitutions with different reactants. The goal was to compare different reaction mechanisms and connect them to the energetic evolution. While the focus lied on the reactions, the chemical bond still played an important part as specific bonds were supposed to be broken and formed. In study I, the students were asked to think aloud while interacting with the environment. The instructor did intervene as little as possible and spoke only to remind the students to think-aloud.

That being said, there were some technical difficulties that the instructor had to fix as this study was performed on a pilot version of the software.

In study II and III, the students were presented with open-ended questions on the instructions. The questions were supposed to elicit contrasting and comparing, as well as reflection. Examples of these questions are "What differences or similarities did you notice?", "Are your conclusions coherent with what you observe?", "What changed?". They were answered in writing directly onto the instructions. Here, although not explicitly named, the focus lied on the potential energy, the force that they might have experienced and the reaction mechanisms. Still, we found many notions of the chemical bond that are re-analyzed in this chapter.

8.1.3 Design

A summary of the study designs is given in Figure 8.2. The data analyzed stems in all three cases from a problem-solving phase. In the case of study II and III, this was followed by an instruction phase, a design commonly known as PS-I.[9,16,17,19] Typically, students are presented with a problem that they do not have the sufficient skills for to solve.[18] This leads the students to create inaccurate solutions and by then discovering the inaccuracy of them, the students get to identify deep features that are salient to the given problem.[17]

While in study I, the focus lied on identifying technical problems of the program and usability issues, in study II and III, the focus lied on learning according to this design principle. To implement this, we chose an exploratory approach. When breaking bonds in the first task, the students became aware of the connection between the potential energy and the molecular structure. Then, as the second and third task follow the same reaction mechanism, the students could apply this connection to contrast, generalize and identify deep features of chemical reactivity – mechanism elucidation as minimal energy problem, the role of molecular and partial charge and the role of frontier molecular orbitals. As the analyzed data stem from this part of the intervention, we refrain from elaborating on the rest of the design.

8.1.4 Procedure

The procedure of the problem-solving phase was close to identical for all three studies. The only difference lied in the qualitative data obtained. For study I, the students thought aloud during the process while in studies II and III, they answered open-ended questions in written form directly onto the instruction sheet. The quantitative data obtained in each study differed based on the goals of each study, the details can be

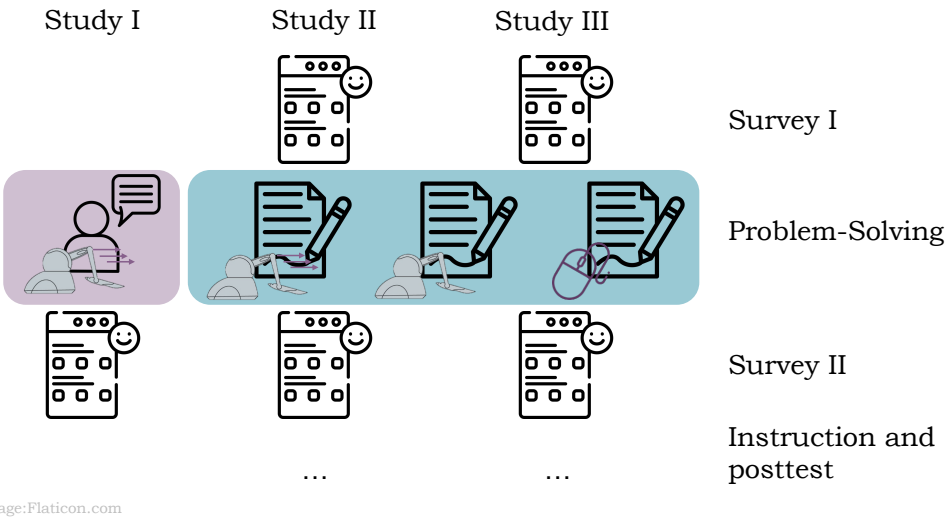


Figure 8.2: Overview of the designs of the studies providing the data analyzed in this chapter. Study I only employed one condition (purple) while study II and III compared students who received haptic feedback (indicated by arrows) with students who used the computer mouse or who used the haptic device without haptic feedback (blue). The qualitative data were obtained during the problem-solving phase.

gathered from the respective chapters (chapter 4 for study I, chapter 6 for study II and chapter 5 for study III).

Consent was sought from all students before the start of the studies. After a first phase of quantitative measures, the students performed the problem-solving phase. First, the students were given familiarization tasks that focused on rotation, zooming and dragging and dropping the atoms. After self-proclaimed proficiency, the students moved on to the learning tasks. These three tasks consisted of i) a single molecule that they could manipulate as they pleased and ii) two nucleophilic substitutions with differing reactants. In addition to exploring different reaction mechanisms, the students could explore the effect of molecular charge or visualize orbitals and electronic density. After finishing the tasks, the students again filled out questionnaires for the quantitative measures before receiving the instruction phase (for study II and III) or being finished (for study I). All studies took between 90 min and 120 min. During the whole process, the instructor aimed to be as little involved as possible. Still, there were some technical issues and the students sometimes struggled with the three dimensional navigation of the haptic device, with which the instructor then assisted.

8.1.5 Analysis

As we were interested in the complexity of views that students bring to and have of the learning experience, the resulting qualitative data were analyzed using an experiential realist, constructionist thematic analysis approach with a semantic orientation.[215,223] Specifically, we assumed that the students' knowledge of the world emerges from active experiences in the world. "[T]ruth is not external to the human mind. Rather, truth is the product of the human mind" (p. 230 of Ref. 223). This is in contrast with the postpositivist view that there is one true learning trajectory or at least a set of probable learning processes to be uncovered.[224] However, it should be noted that I, who conducted the analysis initially, entered into the analysis with a postpositivist perspective, but found that this perspective was restrictive in the analysis and that to truly capture the perspectives and hence realities of the students, a shift towards a constructionist stance was needed. I conducted the analysis alone and ensured reliability through detailed reflective notes on the process as well as feedback from peers and supervisors.[215]

In the following, we briefly outline the analysis process of the studies. In all studies, we started with certain a priori codes that we expected to find in the data based on the specific conceptual framework of the study. We then coded the data in a fine-grained manner, developing codes as needed and allowing an unlimited number of codes per segment in order to avoid missing any inductive meanings. This process was continued until the full data set was coded and no further codes were identified. The codes were then grouped into topic summaries from which themes were derived.

In study I, we focused on usability of the environment and whether or not the environment was interpreted as intended. The a priori themes of study II and III were grounded in the theoretical framework of embodied learning[3] and focused on how the haptic feedback was perceived and whether the visual and haptic representations were connected and interpreted by the students as intended. In addition, we were also interested in inductive themes around the interpretation of the learning environment.

In all three studies, we found that the students brought a wide variety of conceptions concerning the visual representation of the molecules. In particular, we found that the visual representation of the chemical bond as sticks that were displayed depending on interatomic distances could confuse the students and even make them distrust the simulation. As we found this in all studies although the focus of the analysis was elsewhere, we decided to explore further and thus combined the topic summaries around chemical bond conceptions and re-analyzed them. For this iteration, we were interested in what conceptions of the chemical bond the students might have activated or challenged in the learning environment and how they might have been influenced by the chosen

visual representation. We expected known conceptions such as the different kinds of bonds, ionic, covalent and metallic, possibly an energetic understanding that breaking and forming bonds might be related to energy requirement and release.[23] As we were again following a thematic analysis approach, we were open for further conceptions. All fragments previously coded with a chemical bond related code were recoded according to the model that they activated. These codes were then organized into themes that are presented in section 8.2.

For reporting, the answers were translated from German to English including possible mistakes.

8.2 Results

The final codes that were organized into themes are presented in Table 8.2. We generated three themes that capture how the students conceptualized the chemical bond, i) the chemical bond as rigid entity, ii) the chemical bond related to energy and iii) orbitals as indicators of the presence or stability of the bond. First semester students in particular, but even students of other semesters, tended to emphasized the capability of an atom to form a bond and the kind of bond that might be formed, hence perceiving the chemical bond as rigid entity (see Figure 8.3).

Table 8.2: Themes and specific codes.

Theme	Codes	Description
The chemical bond as rigid entity	Ability to form bond Bonding as spontaneous Different kinds of bonds Expectation of event at bond breaking or formation	Breaking and forming of a bond is a sudden event. The bond is either there or not, it is a physical being that is created and destroyed. To do that, the two atoms need to have the ability to form this physical entity.
The chemical bond related to energy	Activation energy Bond as minimal energy distance Bonding lowers energy Other relation of energy to bonding	Breaking and forming of a bond is a process of moving towards a minimal-energy structure. Hence, the end product of the bond reduces the total energy of the molecular system. That being said, an activation barrier needs to be overcome to arrive at the product.
Orbitals as indicators of the presence or stability of the bond	HOMO = unstable bond No orbital, no bond Orbitals separation is bad	As orbitals describe electrons and electrons are the substance of a chemical bond, the orbitals describe the bond now. They can therefore also indicate stability of a molecule. Close to all conceptions found in this theme were inaccurate.

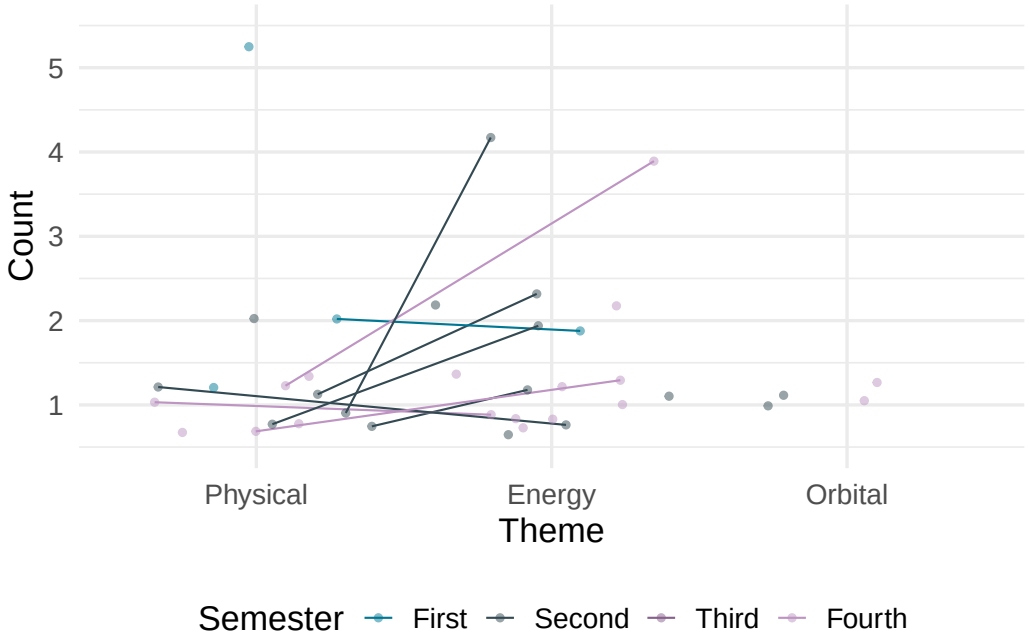


Figure 8.3: Count of codes assigned to the generated themes. This is to give an impression of the distribution of the different utterances and written answers but should by no means be interpreted as statistical evidence. Individual students are represented by a point if they were only situated in one theme or as a line, if they used more than one model throughout the intervention.

For example, J. A. of the first semester asks, whether "hydrogen actually [is] able to form this bond". Some of these observations were directly activated by the choice of visualization such as "Bond extends until it breaks." (C. X., second semester) or "I have the feeling that I can pull both atoms equally far away before the bond breaks" (E. E., second semester). Some students even showed confusion when nothing specific happened or changed upon the event of formation or breaking of a chemical bond.

> So, I am doing this and then, when it forms a bond here, absolutely no difference. (N. Y., first semester)

> E rises the closer I move Cl to C even before + after bonding. (M. M., fourth-semester; E refers to the energy graph)

We attribute this to a certain extent to the ball-and-stick representation of the molecular system. As breaking a stick in the perceivable world means a sudden event, the students might have translated this characteristic to the chemical world. In addition, this might be supported by other binary language commonly employed in chemistry, such as the *breaking* of a bond or the *formation* of a bond. To a chemist, it is clear that this *breaking* implies a process, but in colloquial language, *breaking* is associated

with a sharp, brief and quick event. The second theme 'the chemical bond related to energy' captures conceptions such as connecting an increase or decline of energy to the molecular structure or the chemical bond as a minimal-energy distance. This theme allows chemical bonding to be a process instead of a sudden event. However, some students articulated their views in both themes at the same time, for example when they expected a sudden change in energy at the point of bond formation (see excerpt by N. Y. above, the expectation is made explicit by the word *even*). This blending of themes is in agreement with the work by Becker and Cooper[191] who found that science students conceptualized potential energy mainly as capability for a structure to change, stored energy or as related to stability. Hence, if a student conceptualized potential energy as capability (for bond formation) and the chemical bond as a rigid entity in which energy is stored, the physical entity and energetical understanding of the chemical bond will not necessarily contradict each other.

Examples of this theme include

> As a bond appears, the energy falls again until reaching a minimum at a specific distance. (A. T., fourth-semester)
>
> Energy increases with distance of Br^- & decreases when Cl^- bonds. (C. M., second semester)
>
> The bonding enthalpy is released when a bond is formed. (M. K., fourth semester)
>
> If I come closer it raises until it ends up bonding and stabilizing. (F. P., fourth semester)

Finally, a few students conceptualized the 'chemical bond in connection to orbitals'. Although only found in three students and potentially activated in a biased way as we asked them explicitly to actively explore the orbital visualization feature of the environment, we find it worth mentioning as it illustrates the difficulties moving between an 'electron as particle' view and an 'electron as described by a wavefunction' view. These students understood that their conception of a chemical bond is potentially incomplete in the given context as they could not anymore rely on a localized electron. As a consequence, they tried to connect the new information – the electron is described by a wavefunction or orbital – and the known information – a bond is where two electrons are between atoms.

> Yes, because ipso to the amino group exists no LUMO [lowest unoccupied molecular orbital], therefore there cannot be a bond formed there. (B.E., second-semester)

In this example, B.E. expects that there needs to be some space into which the electron has to be able to move. Orbitals as regions in space for electrons to go into as opposed

to mathematical models describing the electron is a known misconception.[26] As the concept 'electron' changes so much during chemistry education, it is of high importance to consolidate related concepts such as the chemical bond accordingly.

In general, we found that some students only employed a single model. Others employed one model over another at different stages of the problem-solving. Finally, there were students who employed multiple models in one single utterance, indicating potentially an attempted integration of the models. This highlights the unawareness of the distinctiveness between the models and of the context-dependence of them. Although not representative due to the limited amount of data, there is an indication that the students start with mainly the physical entity conception and subsequently learn about the molecule as minimal-energy structure. First, they struggle, as they try to integrate this new information into the chemical bond as physical entity conception – indicated by the multiple second-semester utterances that use both models at the same time.

> Both graphs show a peak in energy prior to the breaking of the band. Chloride seems to leave more easily. (S. X., second-semester)

> If a bond is formed the change is insignificant. [...] The bump is because the Br-C Bond is cleaved and the steep slope because Cl-C Bond is formed. (I. D., second-semester)

S. X. indicates with an event (the peak) *prior* to the breaking of the bond that the breaking itself is an event as well. An indication of timing is only meaningful if bond breaking is an event as opposed to a process. However, they further mention *leaving*, a verb that indicates a process. Similarly, in the second quote, *bump* denotes a sudden event at which the bond is broken. However, formation of the bond is described as a process connected to the slope in energy. Finally, in the fourth-semester, they start to learn when to apply which model.

8.3 Discussion

Thematic analysis of the data of three studies all investigating the effect of haptic feedback on learning quantum chemical concepts resulted in the common theme that chemical bonding is conceptualized in various models. Here, we re-analyzed the complete data and generated three themes that describe how students conceptualized the chemical bond in this particular context.

We found that chemical bonding was mostly viewed as an event ('the chemical bond as rigid entity') by 15/31 students and as a process ('the chemical bond related to energy') by 19/31 students with 9 students using both. The third theme 'Orbitals as indicators of the presence or stability of the bond' was found substantially less frequently in only

four students (see Figure 8.3). However, these conceptions were not mutually exclusive and could be activated at the same time. Arguing in the framework of p-prims as introduced by diSessa,[74] the former conception is grounded in intuitive physics in the form of a metaphorical breaking of sticks. However, we cannot simply identify the process conception as the more proficient conception. Professional chemists are known to employ a diverse set of models when talking about the chemical bond.[22] Scientific models are introduced to explain a phenomenon and in the highly diverse field that chemistry is, it makes sense that these models vary. From a philosophical perspective, Hendry[242] has come to a similar conclusion when analyzing the conceptions of the chemical bond in a historical context. He identified i) the structural conception, in which "chemical bonds are, at least for molecular substances, material parts of the molecule that are responsible for spatially localized submolecular relationships between individual atomic centers" (p. 917) and ii) the energetic conception, according to which "facts about chemical bonding are just facts about energy changes between molecular or supermolecular states" (p. 919). While the latter is more agnostic and better compatible with quantum chemistry, it implies explanatory loss, a characteristic highly important in experimental domains.

The third and less prominent theme that we generated related the newly learned information about orbitals to the prior knowledge of the students. Defining the chemical bond in terms of superposition of valence atomic orbitals is found in literature.[243] However, here, we mainly found inaccurate attempts to connect these concepts, such as the idea that an absence of a specific orbital indicates the absence of a bond or that the presence of the highest occupied molecular orbital at a specific bond indicates an instability at this location (see Table 8.2). Again, we do not identify the model as the source of the difficulty, but rather the inability of the students to successfully integrate it into their prior knowledge as a model that is distinct from the others but equally important in the correct context.

Although most of the students did experience forces, we did not find a fourth conception of the bond often found in literature, namely the definition of the chemical bond as the forces that stabilize the molecular structure.[244] It was even argued that a force-based approach might be best to describe the chemical bond in an educational context.[22,199,245,246] Here, the bonds were visualized based on interatomic distances and hence appeared not only when the structure was stable. This conception could be supported by not visualizing the bonds at all or only if no resulting force is present.

The plurality of models becomes only problematic when the students try to integrate them to a single model, as we have found on multiple occasions. From initial chemistry learning on, the students are confronted with this variety of models. With this analysis,

we are not judging the diverse models employed by the students or the instructors, but imply that the introduction to this diversity without limiting or confusing the students is crucial for successful higher chemistry education.

As we discuss in more detail in chapter 2, many models are of metaphorical nature. Del Re called models *idealized experiments* in the sense that they can successfully predict measurable outcomes based on a metaphorical model system.[44] Many misconceptions that we mentioned in our conceptual framework can be explained by incorrect blending, combining or overgeneralizing of these metaphorical model systems. For example, the students in Boo's study applied the metaphorical model system of physical object to the concept of energy requirement, leading them to inaccurately assuming that as constructing a physical object requires energy, so does bond formation.[200] Similarly, visualizations of orbitals as physical spaces might potentially create the inaccurate metaphorical mapping of orbitals as a space where electrons are or where electrons move into.[26]

8.4 Conclusion

8.4.1 Implications

Students are taught a number of different models in their higher chemical education. While chemists focus on the models helpful to their specific field, such as hybridization and the octet rule for organic chemists or one-electron wavefunctions and potential energy hypersurfaces for quantum chemists, the students are confronted with them concurrently. In that process, we suggest that it is crucial to make the students aware of the different models of the same concept and what the different strength and applications of each are.

One way to support the students in their learning is through experiential learning. The real-time interaction with molecules and experience of the forces that would act on that selected atom that was offered in the discussed studies lead the students to challenge existing conceptions of fundamental quantum chemical concepts generally, such as the Born-Oppenheimer potential energy hypersurface or minimal energy structures. This experiential quantum chemistry learning warrants further investigation.

8.4.2 Limitations

The data for this analysis were limited in the sense that conceptions of the chemical bond were not the primary focus of the data collection and hence, questions focused on the connections between the energy graph and the simulation and the students were not explicitly asked about their interpretation of the visual representation of the molecules.

Accordingly, our data are limited to events where the students were confused or in disagreement with the visual representation and mentioned this on their questionnaire or in their think-aloud interview. Hence, it would be interesting to discuss these conceptions in more depths with the students to be able to obtain a better understanding of the variety and potential hierarchy of these conceptions. Furthermore, the studies were framed as a testing of the learning environment for future generations of chemistry students. We could therefore expect certain distrust in the representations including the presence of the chemical bond. While here, the chemical bond as a construct that is there or not is highly supported by the visualization, future research should explore how other visualizations, such as the electron density, influence the activated conceptions or how the students would reason about the chemical bond without visual input at all.

9

Conclusions and Outlook

In this thesis, we explored the role of conceptual metaphor for quantum chemistry learning. We argued that as chemistry generally and quantum chemistry in particular employ many in-experiential concepts, the explicit connection of these concepts to experiences familiar to the students is of crucial importance. To that end, we designed a learning environment in SCINE Heron[49] that allowed the students to interact with molecules in real-time. Specifically, the students saw the molecules on the computer screen, picked an atom either with a computer mouse or with a haptic device and, pulling on it, felt the repulsion or attraction acting on the selected atom. With the learning environment, we created both grounding and offloading opportunities, two cognitive mechanisms theorized to be crucial for learning by the grounded and embodied learning framework.[3] Furthermore, the design process of the learning environment was inspired by principles from research on multiple[98] and multimodal[186] representations. Chapter 2 derived the theoretical framework and elaborated on the chosen design principles for the learning environment while chapter 3 detailed the existing empirical research on student conceptions as well as previous applications in interactive (quantum) chemistry. In particular, students often overgeneralized the electron-pushing formalism resulting in neglecting of the repulsive forces in chemical reactions.[187] We showed that many misconceptions connected to potential energy and chemical bonds might be prohibited through explicit explanation of the advantages as well as limits of metaphors used to teach them and of implicit metaphors in models. For example, *Energy as a Substance that Can be Stored* is a useful analogy to learn about energy transfer, but neglects energy conservation.[193]

Chapters 4 to 8 describe studies conducted as part of this dissertation. In chapter 4, we described a usability study in the scope of which we tested a pilot version of the environment on undergraduate students. An improved version was applied in three preparation for future learning studies with second- (see chapter 6) and fourth-semester (see chapters 5 and 7) students.

We found that without explicit explanation of the conceptual metaphor, the students were not able to successfully map the embodied experience to the chemical domain. Moreover, the students were distracted by the haptic feedback and – without successfully being able to construct the metaphor – this lead to hindered learning compared to students who could focus on the visual information that was easier to interpret. In a follow-up study to the study described in chapter 5, we explored how making the conceptual metaphor of experiencing the force that acts on a selected atom explicit affects learning and relevant mediators. To that end, we compared students who received haptic feedback together with an explicit explanation of the metaphor to students who received a traditional visualization of force as arrows. We found that explicit metaphors comply with the design principle *affective draw*[19] and that they increase perceived relevance through increased modal priming.[82]

Finally, throughout this work we noticed confusion and to a certain extent fixation of students on the chemical bond concept. For this reason, we triangulated the qualitative data collected in all studies and explored the use of bonding models by the students in an analysis described in chapter 8. We found that the students mainly conceptualized the chemical bond as physical entity or as related to energy. Few attempts to model bonds with orbitals resulted in incorrect conceptions. We argued that the distinction between physical phenomena in chemistry and the models used to describe them needs to be clearer. The role of representational tools such as models is unique in the chemical domain and this representational competence needs to be learned as well.[98] This thesis highlights the importance of truly understanding the students prior conceptual metaphors as well as the implicit metaphors present in applied models, terminology and visualizations in order to be able to transform the former to the latter.[68]

We mentioned the limitations of the individual studies in the respective chapters. To summarize, the two most prominent limitations of this work were the framing of the studies and the recruitment process for studies conducted in the students' spare time, the user study (see chapter 4), the study with second-semester students (see chapter 6) and the posttest that was administered via e-mail for the study described in chapter 7. The studies were framed as a testing of a learning environment for future generations of students. This lead the students to focus on giving feedback and searching for flaws in the design as opposed to learning. Furthermore, the first two years of the bachelor

programs are extremely competitive and the students were reluctant to participate in a study that was not credited towards their studies. To conduct large quantitative learning studies on the target audience, studies should either be incorporated in a course (as we did for the studies described in chapters 5 and 7) or be rewarded with credits (as is already done for psychology students of the University of Zurich, for example).

This work has sparked numerous intriguing questions for future research to address. First, the learning environment has potential to be further improved. For example, allowing the students to feel the activation barrier while only allowing movement along the reaction coordinate might facilitate the controlling of the haptic device while still providing the embodied experience to construct the target conceptual metaphor. This suggestion was brought forward by a student after an intervention. Next to improvements of modality, further target concepts could be explored with similar intervention designs. For example, students often struggle with the atomic or molecular orbital concept.[26] Interactively creating molecular orbitals might facilitate learning about the linear combination of atomic orbitals approach. Furthermore, Heron already now displays molecular orbitals as isosurfaces and offers interactive selection of active spaces as interface to AutoCAS.[247–251] The manual selection process requires high chemical intuition and is thus particularly complex especially for novices. This visual feature might therefore be further valuable for learning about more advanced methods that require active space selection such as for example density matrix renormalization group.[252–264]

Moreover, the distance between the hand either holding the haptic pen or the computer mouse and the molecule on the laptop screen was quite far. In their review, Brockmole and co-workers describe the hand-proximity effect which suggests that the distance between the hand and the object under study is important for visual recall as well as for abstracting commonalities.[265] This distance could be reduced by employment of immersive virtual reality (iVR). iVR has recently gained momentum in the embodied learning community and the higher level of possible immersion might be beneficial for a chemical context.[45,51,82,266,267]

Finally, the role of conceptual metaphor in chemical learning content should be given more attention to. While literature on metaphors in chemistry is rich, the role that this plays in designing learning content in chemistry education is underexplored. Bhushan and Rosenfeld[65] summarized the issue in their work on metaphorical models in chemistry:

> Our own lack of attention to the metaphorical nature of a particular model can blind us as teachers to certain misunderstandings of students. As we use models in teaching, it may be important to recognize that the leading edge and perhaps the most important part of the model for students, is the

> metaphor. Indeed, we might say that where we use models in our teaching, the metaphor is the currency of the teacher-student transaction. (p. 581 of Ref. 65)

This implies a need for understanding what metaphors are implicitly applied in different models, visualizations and terminology as well as a need for understanding how the students integrate these metaphors, which ones are successfully mapped and which ones are frequently mismapped.

A

Concept Inventory

The posttest was developed in-house and applied as well as continuously improved in the usability user study (see chapter 4), the mixed-method learning studies (see chapters 5 and 6) as well as in the follow-up study (see chapter 7). In section A.2, the different versions of the posttest are presented.

A.1 Learning objectives

1. Potential Energy Surface (PES)
    1. Change in molecular structure(s) = change in potential energy
    2. From 1.1 it follows that from all configurations of atomic positions (molecular system in various stretched, compressed, and equilibrium situations), a continuous potential energy (hyper)surface can be created; with every point on this PES corresponding to a specific spatial arrangement of the atoms.
    3. Local minima on this surface correspond to stable molecules.
    4. To stretch a bond corresponds to a displacement on this high-dimensional surface in a specific direction.
    5. From 1.4 it follows that if the structure from which a bond is broken is corresponding to a minimum on the PES, the (potential) energy of the system will increase during this process of bond breaking (i.e. energy will be required to break the bond).
    6. Saddle points on this surface correspond to transition state structures.

7. A reaction coordinate corresponds to a (minimum-energy) path on this PES from one minimum to another passing a saddle point (transition state).

2. Force

    1. The values of the force components experienced by an atomic nucleus are determined by the components of the slope of the PES in the direction of the coordinates of that nucleus and at the point i.e. the current spatial arrangement.

    2. From 2.1 it follows that the force operating on every atom is 0, if the structure represents a stable molecule, i.e., if it corresponds to a minimum on the PES.

    3. From 2.1 it follows that the force operating on an atom while moving along a reaction coordinate will be opposite to the direction of reaction progress along the reaction coordinate until the saddle point is reached (i.e., when the activation barrier is overcome), after which it will pull one into the product valley (negative value of a negative slope yield a push into direction of reaction progress).

3. Reactivity

    1. From 1.2 it follows that there will exist, in general, more than one way to connect two minima on the PES or in other words more than one possible mechanism from the reactants to the product. However, one reaction path is favourable which is usually the one with lowest activation barrier, i.e., with lowest-lying a saddle point.

    2. From 1.2 it follows that there is usually more than one way for the structures to change, i.e. that there is often more than one possible reaction product. However, one reaction path is often (thermodynamically or kinetically) favourable (the one with the energetically lower lying products for comparable activation energies or the one with the lowest barrier or energy span, respectively), the other products are so-called byproducts.

4. Application: S_N2 mechanism

    1. There is one specific mechanism: attack by the nucleophile under formation of a stable intermediate, dissociation of the leaving group.

    2. The nature of the nucleophile and leaving group matters because of differing bond strength (slope of PES along the reaction coordinate that breaks the bond).

A.2 Questionnaire

A.2.1 Usability Study

We first present the initial questionnaire as developed for the usability study presented in chapter 4. Second, we present changes made for the subsequent studies with explanations.

Note: correct answers are marked with (∗), every item corresponds to a learning objective from the previous section.

1. What action will keep the potential energy constant?
    1. A change from the reactants to the product.
    2. A change of the location of the molecule in a vacuum. (∗)
    3. A change of the bond angles in a molecule.
    4. A change of distances between atoms or molecules.

2. What do individual points on a potential energy graph correspond to?
    1. Different spatial arrangements of atoms. (∗)
    2. Different levels of excitation.
    3. Different vibrational frequencies.
    4. Different atomic velocities.

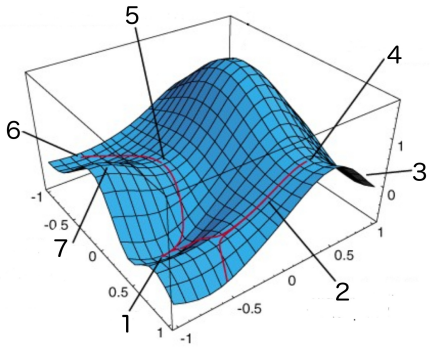
3. Which of the following statements is **true** for the potential energy of a system that has reached the thermodynamically most stable form?
    1. It has reached its maximum, since the kinetic energy will be minimized.
    2. It has reached its maximum, since the kinetic energy will be 0.
    3. It has reached a minimum, since energy will be needed to change this optimal conformation. (∗)
    4. More information is needed to comment on the potential energy.

4. Which of the following is **false**?
    1. Forming a bond will change the potential energy.
    2. Stretching a bond requires energy.
    3. Forming a bond releases kinetic energy.
    4. Stretching a bond releases energy. (∗)

5. If a random bond in a stable molecule is being broken, how will the potential energy of the system change?

    1. It will increase. (∗)
    2. It will decrease.
    3. If it decreases or increases will depend on how much energy is stored in the bond.
    4. It will neither increase nor decrease.

6. Which of the following points could correspond to transition state structures?



Adapted with permission from H. B. Schlegel, *WIREs Comput. Mol. Sci.* **2011**, 1, 790-809. Copyright 2011 John Wiley & Sons.

    1. 1, 3, 6
    2. 2, 7
    3. 4, 5 (∗)
    4. None

7. A reaction coordinate on a potential energy surface is …

    (Hint: A reaction coordinate describes a reaction and therefore requires the conversion of kinetic energy into potential energy to climb the surface.)

    1. … the minimum-energy path on the potential energy surface between some reactants and products. (∗)
    2. … any path on the potential energy surface from reactants to products.
    3. … always the shortest path, i.e., a straight line connecting reactants and products.
    4. … is a single direction on the potential energy surface that corresponds to the stretching of a bond.

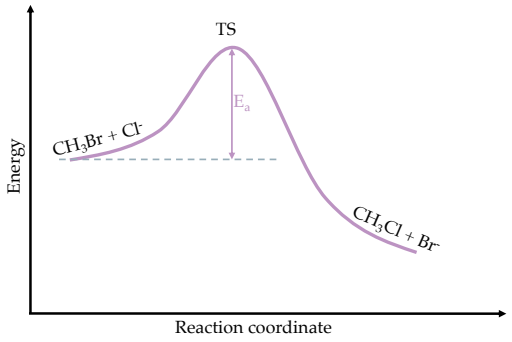
8. How does the potential energy relate to the force which acts on an atom?

    1. If the potential energy increases, a force will act on an atom in the same direction as the structural change.
    2. If the potential energy increases, a force will act on an atom opposite to the direction of structural change. (∗)
    3. Energy and force are two separate concepts that do not relate.
    4. There are only attractive forces present during a reaction.

9. What is the resulting force that acts on an atom in a stable molecule?

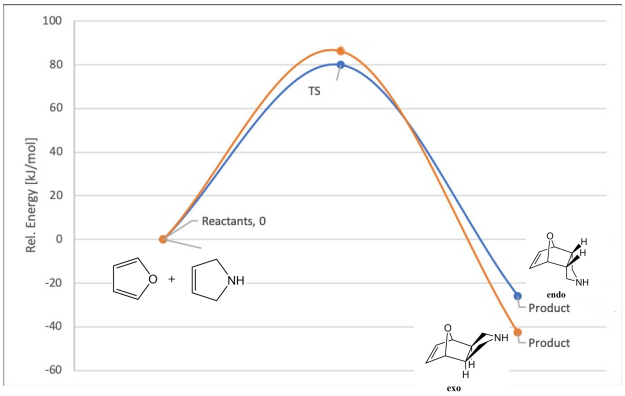
    1. This depends on the mass of the atom.
    2. This depends on the potential energy.
    3. This depends on the kinetic energy.
    4. None (∗)

10. Explain how the following image relates to forces present in the system.



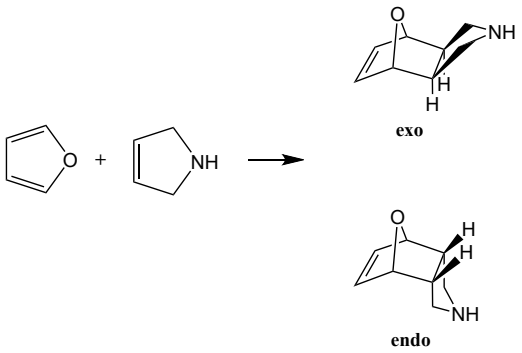
    1. Since the energy of the reactants is higher than the one of the products, the force between the educts is attractive and reaction will occur.
    2. First, the force between the reactants is attractive as long as the energy goes up, and will then become repulsive.
    3. First, the force between reactants is repulsive as long as the energy goes up, and will then become attractive. (∗)
    4. No conclusion about the force can be made, since this is an energy diagram.

11. Given the following energy diagram, which reaction is thermodynamically more likely to occur given enough time? (TS = Transition State)



1. The reaction resulting in the endo product (blue), since the energy of the product is higher.
2. The reaction resulting in the endo product (blue), since the activation barrier is lower.
3. The reaction resulting in the exo product (orange), since the product is energetically lower. (∗)
4. The reaction resulting in the exo product (orange), since the activation barrier is higher.

12. In the following reaction, two products are formed. Which of the following statements is **wrong**?



1. If given a long time, the thermodynamic product will be formed, which is the one with the highest negative energy difference between reactants and product.
2. If given a short time, the kinetic product is formed, which is the one with the lowest activation barrier.
3. The structural difference of the products is determined by how the reactants are spatially oriented with respect to one another at the start of the reaction.
4. Only one product will be formed and we can determine which one by looking

at the kinetic as well as the thermodynamic results (such as relative energies of products and activation barrier). (∗)

A.2.2 Learning Studies 2022

Table A.1: Analysis of performance in the usability study per item. Items that were changed are indicated in **bold**.

Item	*Mean*	*SD*	*Skew*	*Item Difficulty*	*Item Discrimination*	*α if deleted*
Q1	0.77	0.44	-1.45	0.77	0.51	0.80
Q2	0.38	0.51	0.54	0.38	0.33	0.81
Q3	0.69	0.48	-0.95	0.69	0.68	0.78
Q4	0.77	0.44	-1.45	0.77	0.72	0.78
Q5	0.62	0.51	-0.54	0.62	0.61	0.79
Q6	0.62	0.51	-0.54	0.62	0.31	0.81
Q7	0.31	0.48	0.95	0.31	0.61	0.79
Q8	0.92	0.28	-3.61	0.92	0.18	0.82
Q9	0.15	0.38	2.18	0.15	0.27	0.81
Q10	0.46	0.52	0.18	0.46	0.44	0.80
Q11	0.62	0.51	-0.54	0.62	0.37	0.81
Q12	0.46	0.52	0.18	0.46	0.50	0.80
					Mean inter-item-correlation = 0.261, Cronbach's α = 0.814	

We analyzed the items based on their item difficulty and item discrimination index. The item difficulty describes the proportion of participants who correctly answered the question with a low index indicating high difficulty and high index indicating low difficulty; the item discrimination index describes how well the item performance predicts overall performance ranging between -1 and 1 with -1 indicating a negative relationship between item performance and overall performance and 1 indicating a perfect relationship between item performance and overall performance.[268,269] The statistical analysis is presented in Table A.1. Changes are indicated in *italic*.

Item 4 was changed as follows to increase precision of the terminology used.

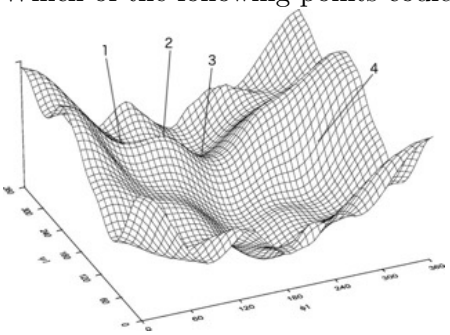
Which of the following is **false**?

1. Forming a bond will change the potential energy.

2. Stretching a bond requires energy.
3. Forming a bond releases *vibrational* energy.
4. Stretching a bond releases energy. (∗)

Item 6 was changed as the identical potential energy surface was displayed in the instruction material of the course.

Which of the following points could correspond to transition state structures?



Adapted with permission from G. N. Zamarbide et al., J. Mol. Struct. THEOCHEM. ***2003****, 666-667, 599-608. Copyright 2003 Elsevier B.V.*

1. 2, 3
2. 1, 4
3. 1, 3 (∗)
4. 3, 4

Item 8 was changed as follows to introduce the misconceptions that force is proportional to the energy instead of the change in energy and that the force is operating between molecules only as distractors. This was supposed to increase the difficulty of the item as it was correctly solved by a lot of students (item difficulty 0.92).

How does the potential energy relate to the force which acts on an atom?

1. If the potential energy increases, a force will act on an atom in the same direction as the structural change.
2. If the potential energy increases, a force will act on an atom opposite to the direction of structural change. (∗)
3. *The potential energy is proportional to the force.*
4. *The force is related to two molecules while the potential energy is an attribute to one only.*

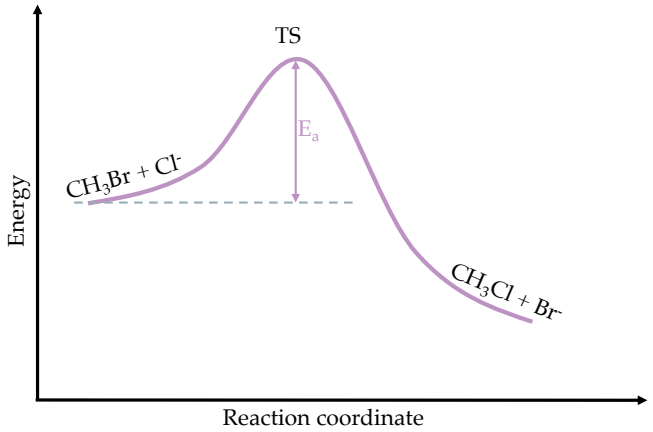
Item 9 was changed due to its low item difficulty (0.15) indicating that the item was quite difficult. Hence, the phrasing of the question was adapted as follows.

*On what quantities does the resulting force that acts on an atom in a **stable** molecule depend?*

1. This depends on the mass of the atom.
2. This depends on the potential energy.
3. This depends on the kinetic energy.
4. None (∗)

Item 10 was changed to increase precision in terminology.

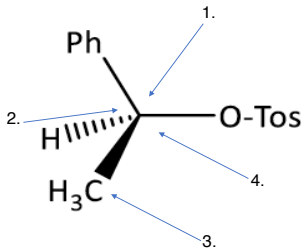
Explain how the following *sketch* relates to forces present in the system.



1. Since the energy of the reactants is higher than the one of the products, the force between the educts is attractive and reaction will occur.
2. First, the force between the reactants is attractive as long as the energy goes up, and will then become repulsive.
3. First, the force between reactants is repulsive as long as the energy goes up, and will then become attractive. (∗)
4. No conclusion about the force can be made, since this is an energy diagram.

Item 12 was replaced by the following item since i) the exact reaction from the usability study item 12 was introduced in the course and ii) the wording of the task was quite long.

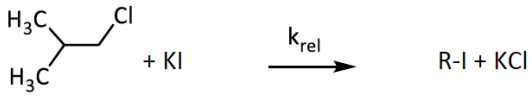
Which attack point is most likely to trigger a reaction?



1. 1.
2. 2. (∗)
3. 3.
4. 4.

For the study presented in chapter 6, two additional items were implemented. The following item was directly taken from the lecture material and hence an isomorphic example.

Why does the following reaction hardly ever take place?

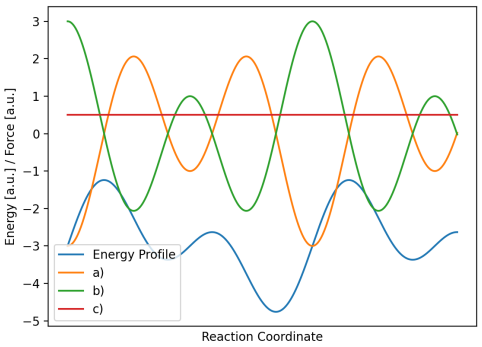


1. Steric hindrance by chloride.
2. Bad electrophile.
3. Steric hindrance by methyl group(s). (∗)
4. Bad nucleophile.

The following item was introduced to test if the mathematical relation between force and energy was understood. We found that the previous version of the posttest was missing an item that tested this knowledge as removed as possible from chemical content.

In blue, an energy profile of a reaction is given which shows multiple transition states and intermediates. Only considering thermodynamic aspects, which of the other curves (if any) is showing the profile of the value of the force operating in the direction of the reaction coordinate (i.e. the direction of the movement of the reaction)?

Hint: The force can be interpreted as a vector that points toward the direction of the reaction coordinate, if the value is positive or away from it, if the value is negative. The value then indicates the length or in our case, the strength of the force.



1. a)
2. b) (∗)
3. c)
4. Not clear from the information given.

A.2.3 Learning Study 2023

In the study of 2023 described in chapter 7, we added per item the open-ended question "Please explain your choice briefly". The statistical analysis is presented in Table A.2. While the answering of the multiple-choice items was enforced, these items were not. In the following, changes from the version applied in the learning studies of 2022 are explained.

Item 1 was changed due to its low item difficulty (0.94) and item discrimination (0.16). Symmetric manipulation and rotation as opposed to translation were introduced as distractors.

What action will keep the potential energy constant?

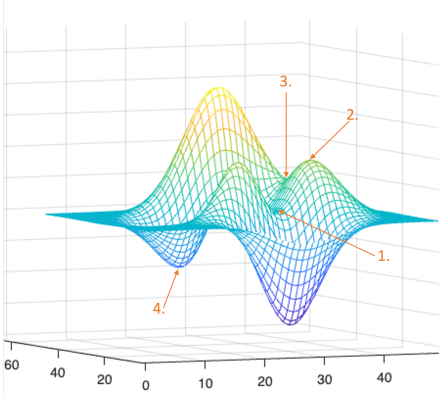
1. A change from *transition state to intermediate.*
2. A *rotation of the molecule around a given axis* (∗)
3. A change of the bond angles in a molecule.
4. A change of *all bond lengths equally.*

Table A.2: Analysis of performance in the learning study 2022 per item. Items that were changed are indicated in **bold**. The items are ordered as they were posed in the questionnaire.

Item	*Mean*	*SD*	*Skew*	*Item Difficulty*	*Item Discrimination*	*α if deleted*
Q2	0.63	0.49	-0.56	0.63	0.42	0.4
Q7	0.38	0.49	0.5	0.38	0.06	0.51
Q8	0.63	0.49	-0.56	0.63	0.27	0.45
Q1	0.94	0.24	-3.84	0.94	0.16	0.48
Q11	0.65	0.48	-0.63	0.65	0.38	0.41
Q4	0.63	0.49	-0.56	0.63	0.19	0.47
Q3	0.88	0.32	-2.43	0.88	0.35	0.44
Q5	0.69	0.47	-0.85	0.69	0.24	0.46
Q9	0.18	0.38	1.74	0.18	-0.04	0.53
Q10	0.72	0.45	-1.01	0.72	0.15	0.49
Q6	0.66	0.48	-0.7	0.66	-0.08	0.55
Q12	0.72	0.45	-0.98	0.72	0.26	0.46
Mean inter-item-correlation = 0.077, Cronbach's α = 0.496						

Item 6 was changed due to its negative item discrimination (-0.08) indicating that item performance did not predict overall performance. The choice of points was improved to more attractive options.

Which of the following points could correspond to transition state structures?



1. 2, 3
2. 1, 4
3. 1, 3 (∗)

4. 3, 4

Item 7 was shortened as follows due to its low item discrimination (0.06).

A reaction coordinate on a potential energy surface is …

(Hint: A reaction coordinate describes a reaction and therefore requires the conversion of kinetic energy into potential energy to climb the surface.)

1. … the minimum-energy path from reactants to products. (∗)
2. … any path from reactants to products.
3. … *a straight line from* reactants to products.
4. … a single direction that corresponds to the stretching of a bond.

Item 9 was changed due to its high difficulty (0.18) and low item discrimination (-0.08). The answer options were adapted in order to be more equal in length and also the (incorrect) proportional relationship between energy and force was further highlighted.

In a ***stable*** *molecule, what force acts on an atom?*

1. *A force proportional to* the mass of the atom.
2. *A force proportional to* the potential energy.
3. *A force proportional to* the kinetic energy.
4. No *force at all.* (∗)

B

Coding Books

The qualitative analyses described in chapters 4 and 5 both followed a thematic analysis approach.[215] Therefore, the data were coded in a fine-grained manner in order to describe them as detailed as possible. From the codes, themes were generated that describe the overarching meaning assigned to the codes. In this appendix, the full coding books from both analyses are listed.

B.1 Usability study

In the usability study described in chapter 4, we focused on technical feedback from the students as well as whether they interpreted the environment as intended. We found that the students mainly struggled with the three-dimensional nature of the molecule representations and that they were not often able to connect all three representations, the molecular simulation, the energy graph and the haptic feedback. The themes were generated from the codes presented in Table B.1.

B.2 Learning study

In the learning study described in chapter 5, the students answered open-ended questions directly unto the instructions in writing. We analyzed the data with the intent to explain the quantitative results that contradicted our hypotheses. The themes were generated from codes listed in Table B.2.

Table B.1: Final coding book of the analysis of the think-aloud data produced in the usability study.

Name	Description	Cases	References
focus			
connect all	Connect all three representations in one line of reasoning.	6	14
connect representation to prior knowledge	That is, prior knowledge that is not represented in any way in the learning environment (i.e. not energy, partial charges while on, force).	6	12
connect visual_mol haptic	Connect molecules to haptic rep.	9	47
connect visual_mol visual_en	Connect molecules to energy graph.	7	15
haptic	Focus on haptic representation only.	3	5
haptic visual_en	Connect haptic rep. to energy graph.	2	5
programming	Comment on implementation.	6	21
visual_en	Focus on energy graph only.	4	4
visual_mol	Focus on visual representation of the molecule only (as opposed to visual representation of graph or haptic representation or the connection between any of them).	6	23
interpretation			
colors as element	Correct assignment of color of ball to chemical element.	2	2
not colors as elements	Incorrect assignment of color of ball to chemical element.	2	3
model	Inability to realize what is a model and what is actually happening. This includes “does this really take place” or “are this many bonds really possible”.	9	47
sensory ability			
not force	i.e. not able to either feel anything, or being able to distinguish attraction and repulsion.	5	13

Table B.2: Final coding book of the analysis of the written answers to open-ended quetsions posed in the problem-solving phase of the PS-I study.

Name	Description	Cases	References
Focus		45	288
connect graph - atom movement	Only introduced when coding non-haptic data; used to code connections between the graph and atomic movement.	19	28
connect graph - haptic	Connection between energy graph and hapticfeedback.	13	23
focus graph	Focus only on energy graph.	43	170
focus haptic	Focus only on haptic feedback.	15	41
focus mol	Focus on molecular representation. Particularly the movement of the atoms.	16	26
Model		45	579
model bond binary	Bond is either there or not, something happens when a bond forms or is broken (e.g. energy spikes). Only coded when event is at formation or breaking (not during formation or after formation).	25	53
model charge	Collecting ideas about charge. e.g. no charge leads to repulsion (probably not equal to charge leads to attraction).	41	92
Model Energy		45	381
model causal chain - repulsion, E change	E changes, because there is electronic repulsion.	2	2
model E distance based	As opposed to E dependent on binary formed or broken bonds.	35	77
model energy level	As opposed to e.g. potential well.	15	20
model energy well	Mentioning of minimum.	13	15
model gradual E change	As opposed to code model energy level, here the students recognize a gradual change in energy (maybe but not necessarily connected to a structural change).	43	177
model likelihood based on activation energy	Argue what reaction would be more likely to take place depending on the energy spike in the graph or the repulsive force by the haptic device.	11	16
model likelihood based on reaction energy	Argue that reaction would be more likely to take place depending on reaction energy difference.	12	16
model what determines energy	They should be aware of the Schrodinger equation at this point. However, it might be that they have no idea what the absolute energy is and why it is chemically irrelevant.	28	58

Model Force		25	43
model force = change in energy	The force is proportional to the change of energy.	6	8
model force = energy	Concerning magnitude. Height of energy indicating how strong repulsion (instead of slope of energy).	9	12
model force as sum	The force is related to a sum of interactions.	2	2
model force energy synonym	Explorative code because in cases "less energy required" might mean "less work/force required" due to the colloquial "this required a lot of energy to do".	3	4
model interaction	Force as interaction different elements have different interactions.	5	5
model likelihood based on physical effort	Argue what reaction would be more likely to take place depending on the how exhausting it was to perform the reaction.	10	12
model multidimensionality	Exploratory code - mentioning of multiple reaction pathways, connection 2D plot / surface, byproducts, etc.	4	4
model resonance structures	i.e. orto/para mentioning, expected electronic structures based on organic chemistry principles.	4	6
Programm(ing)		27	64
program atoms moving	Were the students aware that the atoms were moving without manipulation?!	12	29
aware		10	11
unaware		3	3
program trust	Indication that they try to explain unexpected findings with chemical knowledge instead of flaw of program.	16	27
program usability	Indication that person was aware that they were testing a program - potentially a beta version of a program.	5	8

References

[1] Révész, G. *Psychology and art of the blind.*; Longmans, Green, 1950.

[2] Katz, D.; Krueger, L. E. *The world of touch*; Psychology press, 1989.

[3] Nathan, M. J. *Foundations of embodied learning: A paradigm for education*; Routledge, 2022.

[4] Zacharia, Z. C. Examining whether touch sensory feedback is necessary for science learning through experimentation: A literature review of two different lines of research across K-16. *Educ. Res. Rev.* **2015**, *16*, 116–137.

[5] Johnstone, A. H. Why is science difficult to learn? Things are seldom what they seem. *J. Comput. Assist. Learn.* **1991**, *7*, 75–83.

[6] Tümay, H. Reconsidering learning difficulties and misconceptions in chemistry: Emergence in chemistry and its implications for chemical education. *CERP* **2016**, *17*, 229–245.

[7] Klein, U. *Experiments, models, paper tools: Cultures of organic chemistry in the nineteenth century*; Stanford University Press, 2003.

[8] Shapiro, L. *Embodied cognition*; Routledge, 2019.

[9] Schwartz, D. L.; Bransford, J. D. A Time For Telling. *Cogn. Instr.* **1998**, *16*, 475–522.

[10] Minogue, J.; Jones, M. G. Haptics in education: Exploring an untapped sensory modality. *Rev. Educ. Res.* **2006**, *76*, 317–348.

[11] Haag, M. P.; Reiher, M. Real-time quantum chemistry. *Int. J. Quantum Chem.* **2013**, *113*, 8–20.

[12] Haag, M. P.; Vaucher, A. C.; Bosson, M.; Redon, S.; Reiher, M. Interactive chemical reactivity exploration. *ChemPhysChem* **2014**, *15*, 3301–3319.

[13] Vaucher, A. C.; Haag, M. P.; Reiher, M. Real-time feedback from iterative electronic structure calculations. *J. Comput. Chem.* **2016**, *37*, 805–812.

[14] Vaucher, A. C.; Reiher, M. Steering orbital optimization out of local minima and saddle points toward lower energy. *J. Chem. Theory Comput.* **2017**, *13*, 1219–1228.

[15] Marti, K. H.; Reiher, M. Haptic quantum chemistry. *J. Comput. Chem.* **2009**, *30*, 2010–2020.

[16] Kapur, M. Productive failure. *Cogn. Instr.* **2008**, *26*, 379–424.

[17] Loibl, K.; Roll, I.; Rummel, N. Towards a theory of when and how problem solving followed by instruction supports learning. *Educ. Psychol. Rev.* **2016**, *29*, 693–715.

[18] Kapur, M.; Bielaczyc, K. Designing for productive failure. *J. Learn. Sci.* **2012**, *21*, 45–83.

[19] Sinha, T.; Kapur, M. When problem solving followed by instruction works: Evidence for productive failure. *Rev. Educ. Res.* **2021**, *91*, 761–798.

[20] Zohar, A. R.; Levy, S. T. From feeling forces to understanding forces: The impact of bodily engagement on learning in science. *JRST* **2021**, *58*, 1203–1237.

[21] Bosia, F.; Weymuth, T.; Reiher, M. Ultra-fast spectroscopy for high-throughput and interactive quantum chemistry. *Int. J. Quantum Chem.* **2022**, *122*, e26966.

[22] Nahum, T. L.; Mamlok-Naaman, R.; Hofstein, A.; Krajcik, J. Developing a new teaching approach for the chemical bonding concept aligned with current scientific and pedagogical knowledge. *Sci. Educ.* **2007**, *91*, 579–603.

[23] Hunter, K. H.; Rodriguez, J.-M. G.; Becker, N. M. A review of research on the teaching and learning of chemical bonding. *J. Chem. Educ.* **2022**, *99*, 2451–2464.

[24] Primas, H.; Müller-Herold, U. *Elementare Quantenchemie*; Springer-Verlag, 2013.

[25] Tsaparlis, G. Atomic orbitals, molecular orbitals and related concepts: Conceptual difficulties among chemistry students. *Res. Sci. Educ.* **1997**, *27*, 271–287.

[26] Stefani, C.; Tsaparlis, G. Students' levels of explanations, models, and misconceptions in basic quantum chemistry: A phenomenographic study. *JRST* **2009**, *46*, 520–536.

[27] Dangur, V.; Avargil, S.; Peskin, U.; Dori, Y. J. Learning quantum chemistry via a visual-conceptual approach: Students' bidirectional textual and visual understanding. *CERP* **2014**, *15*, 297–310.

[28] Autschbach, J. Orbitals: Some fiction and some facts. *J. Chem. Educ.* **2012**, *89*, 1032–1040.

[29] Jehi, T.; Khan, R.; Dos Santos, H.; Majzoub, N. Effect of COVID-19 outbreak on anxiety among students of higher education: A review of literature. *Curr. Psychol.* **2022**, 1–15.

[30] Vollmer, G. *Mesocosm and objective knowledge*; Springer Netherlands, 1984; pp 69–121.

[31] Lin, Y. I.; Son, J. Y.; Rudd, J. A. Asymmetric translation between multiple representations in chemistry. *Int. J. Sci. Educ,* **2016**, *38*, 644–662.

[32] Simon, H. A. Information-processing theory of human problem solving. *Handbook of learning and cognitive processes* **1978**, *5*, 271–295.

[33] Barsalou, L. W. Grounded cognition. *Annu. Rev. Psychol.* **2008**, *59*, 617–645.

[34] Newell, A.; Shaw, J. C.; Simon, H. A. Report on a general problem solving program. IFIP Congr. 1959; p 64.

[35] Clark, A.; Chalmers, D. The extended mind. *Analysis* **1998**, *58*, 7–19.

[36] Wilson, A. D.; Golonka, S. Embodied cognition is not what you think it is. *Front. Psychol.* **2013**, *4*.

[37] Varela, F. J.; Thompson, E.; Rosch, E. *The embodied mind, revised edition: Cognitive science and human experience*; MIT press, 2017.

[38] Spencer, J. P.; Smith, L. B.; Thelen, E. Tests of a dynamic systems account of the A-not-B error: The influence of prior experience on the spatial memory abilities of two-year-olds. *Child Dev.* **2001**, *72*, 1327–1346.

[39] Spiel, K. The bodies of TEI – Investigating norms and assumptions in the design of embodied interaction. Proc. 15th Conf. TEI. 2021; pp 1–19.

[40] Lakoff, G.; Johnson, M. *Philosophy in the flesh: The embodied mind and its challenge to western thought.*; New York: Basic Books, 1999.

[41] Laszlo, P. Chemistry, knowledge through actions. *HYLE* **2014**, *20*, 93–119.

[42] Janich, P. Wozu Philosophie der Chemie? *Chemie in unserer Zeit* **1994**, *28*, 139–146.

[43] Kekulé, A. Über die Constitution und die Metamorphosen der chemischen Verbindungen und über die chemische Natur des Kohlenstoffs. *Liebigs Ann. Chem.* **1858**, *106*, 129–159.

[44] Del Re, G. Models and analogies in science. *HYLE* **2000**, *6*, 5–15.

[45] Chatain, J.; Kapur, M.; Sumner, R. W. Three perspectives on embodied learning in virtual reality: Opportunities for interaction design. SIGCHI Conf. Hum. Factors Comput. Syst. 2023; pp 1–8.

[46] Rauscher, F. H.; Krauss, R. M.; Chen, Y. Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychol. Sci.* **1996**, *7*, 226–231.

[47] Harle, M.; Towns, M. A review of spatial ability literature, its connection to chemistry, and implications for instruction. *J. Chem. Educ.* **2011**, *88*, 351–360.

[48] Ehrlich, S. B.; Levine, S. C.; Goldin-Meadow, S. The importance of gesture in children's spatial reasoning. *Dev. Psychol.* **2006**, *42*, 1259.

[49] Bensberg, M.; Brandino, G. P.; Can, Y.; Del, M.; Grimmel, S. A.; Mesiti, M.; Müller, C. H.; Steiner, M.; Türtscher, P. L.; Unsleber, J. P.; Weberndorfer, M.; Weymuth, T.; Reiher, M. qcscine/heron: Release 1.0.0. 2022; https://zenodo.org/record/7038388.

[50] Nathan, M. J.; Walkington, C. Grounded and embodied mathematical cognition: Promoting mathematical insight and proof using action and language. *CRPI* **2017**, *2*.

[51] Chatain, J.; Ramp, V.; Gashaj, V.; Fayolle, V.; Kapur, M.; Sumner, R. W.; Magnenat, S. Grasping Derivatives: Teaching Mathematics through Embodied Interactions using Tablets and Virtual Reality. Proc. 21st Annu. ACM Interact. Des. Child. Conf. 2022; pp 98–108.

[52] Luisi, P. L. Emergence in chemistry: Chemistry as the embodiment of emergence. *Found. Chem.* **2002**, *4*, 183–200.

[53] Ben-Zvi, R.; Eylon, B.-S.; Silberstein, J. Is an atom of copper malleable? *J. Chem. Educ.* **1986**, *63*, 64.

[54] Griffiths, A. K.; Preston, K. R. Grade-12 students' misconceptions relating to fundamental characteristics of atoms and molecules. *JRST* **1992**, *29*, 611–628.

[55] Lin, H.-S.; Cheng, H.-J.; Lawrenz, F. The assessment of students and teachers' understanding of gas laws. *J. Chem. Educ.* **2000**, *77*, 235.

[56] Nahum, T. L.; Hofstein, A.; Mamlok-Naaman, R.; Ziva, B.-D. Can final examination amplify students' misconceptions in chemistry? *CERP* **2004**, *5*, 301–325.

[57] Demircioglu, G.; Ayas, A.; Demircioglu, H. Conceptual change achieved through a new teaching program on acids and bases. *CERP* **2005**, *6*, 36–51.

[58] Schmidt, H.-J. A label as a hidden persuader: Chemists' neutralization concept. *Int. J. Sci. Educ.* **1991**, *13*, 459–471.

[59] Talanquer, V. Students' predictions about the sensory properties of chemical compounds: Additive versus emergent frameworks. *Sci. Educ.* **2008**, *92*, 96–114.

[60] Niebert, K.; Gropengiesser, H. Understanding starts in the mesocosm: Conceptual metaphor as a framework for external representations in science teaching. *Int. J. Sci. Educ,* **2015**, *37*, 903–933.

[61] Crosland, M. *Historical studies in the language of chemistry*; 1962.

[62] Alborn, T. L. Negotiating notation: Chemical symbols and British society, 1831-1835. *Ann. of science* **1989**, *46*, 437–460.

[63] Knight, D. M. *Science in the romantic era*; Routledge, 2016; pp 237–260.

[64] Mahootian, F. *Philosophy of chemistry: Growth of a new discipline*; Springer, 2015; pp 121–139.

[65] Bhushan, N.; Rosenfeld, S. Metaphorical models in chemistry. *J. Chem. Educ.* **1995**, *72*, 578.

[66] Ashcroft, R. E. The double helix 50 years on: Models, metaphors, and reductionism. *J. Med. Ethics* **2003**, *29*, 63–64.

[67] Lakoff, G.; Johnson, M. *Metaphors we live by*; University of Chicago press, 1980.

[68] Niebert, K.; Marsch, S.; Treagust, D. F. Understanding needs embodiment: A theory-guided reanalysis of the role of metaphors and analogies in understanding science. *Sci. Educ.* **2012**, *96*, 849–877.

[69] Lakoff, G.; Núñez, R. *Where mathematics comes from*; New York: Basic Books, 2000; Vol. 6.

[70] Gilbert, J. K.; Watts, D. M. Concepts, misconceptions and alternative conceptions: Changing perspectives in science education. *Stud. Sci. Educ.* **1983**, *10*, 61–98.

[71] Driver, R.; Easley, J. Pupils and paradigms: A review of literature related to concept development in adolescent science students. *Stud. Sci. Educ.* **1978**, *5*, 61–84.

[72] Smith III, J. P.; diSessa, A. A.; Roschelle, J. Misconceptions reconceived: A constructivist analysis of knowledge in transition. *J. Learn. Sci.* **1994**, *3*, 115–163.

[73] Carey, S.; Spelke, E. *Mapping the mind: Domain specificity in cognition and culture.*; Cambridge University Press, 1994; pp 169–200.

[74] DiSessa, A. A. *Constructivism in the Computer Age*; 1988.

[75] Amin, T. G. Conceptual metaphor meets conceptual change. *Hum. Dev.* **2009**, *52*, 165–197.

[76] Venville, G. J.; Treagust, D. F. Exploring conceptual change in genetics using a multidimensional interpretive framework. *JRST* **1998**, *35*, 1031–1055.

[77] Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. *Molecular biology of the cell*; Garland Science, 2008.

[78] Talanquer, V. Commonsense chemistry: A model for understanding students' alternative conceptions. *J. Chem. Educ.* **2006**, *83*, 811.

[79] Fyfe, E. R.; Nathan, M. J. Making 'concreteness fading' more concrete as a theory of instruction for promoting transfer. *Educ. Rev.* **2019**, *71*, 403–422.

[80] Kokkonen, T.; Schalk, L. One instructional sequence fits all? A conceptual analysis of the applicability of concreteness fading in mathematics, physics, chemistry, and biology Education. *Educ. Psychol. Rev.* **2021**, *33*, 797–821.

[81] Chalmers, D. J. Verbal disputes. *Philos. Rev.* **2011**, *120*, 515–566.

[82] Chatain, J.; Varga, R.; Fayolle, V.; Kapur, M.; Sumner, R. W. Grounding graph theory in embodied concreteness with virtual reality. Proc. 17th Conf. TEI. 2023.

[83] Brookes, D. T.; Etkina, E. Using conceptual metaphor and functional grammar to explore how language used in physics affects student learning. *Phys. Rev. Phys. Educ. Res.* **2007**, *3*, 010105.

[84] Martino, M.; Salvadori, A.; Lazzari, F.; Paoloni, L.; Nandi, S.; Mancini, G.; Barone, V.; Rampino, S. Chemical promenades: Exploring potential-energy surfaces with immersive virtual reality. *J. Comput. Chem.* **2020**, *41*, 1310–1323.

[85] Körner, A.; Topolinski, S.; Strack, F. Routes to embodiment. *Front. Psychol.* **2015**, *6*, 940.

[86] Cowan, N. Working memory underpins cognitive development, learning, and education. *Educ. Psychol. Rev.* **2014**, *26*, 197–223.

[87] Evans, J. S. B.; Stanovich, K. E. Dual-process theories of higher cognition: Advancing the debate. *Perspect. Psychol. Sci.* **2013**, *8*, 223–241.

[88] Nathan, M. J. Rethinking formalisms in formal education. *Educ. Psychol.* **2012**, *47*, 125–148.

[89] Wolski, R.; Jagodziński, P. Virtual laboratory—Using a hand movement recognition system to improve the quality of chemical education. *Br. J. Educ. Technol* **2019**, *50*, 218–231.

[90] Howison, M.; Trninic, D.; Reinholz, D.; Abrahamson, D. The mathematical imagery trainer: From embodied interaction to conceptual learning. SIGCHI Conf. Hum. Factors Comput. Syst. 2011; pp 1989–1998.

[91] Johnson-Glenberg, M. C.; Birchfield, D. A.; Tolentino, L.; Koziupa, T. Collaborative embodied learning in mixed reality motion-capture environments: Two science studies. *J. Educ. Psychol.* **2014**, *106*, 86–104.

[92] Magana, A. J.; Balachandran, S. Unpacking students' conceptualizations through haptic feedback. *J. Comp. Assist. Learn.* **2017**, *33*, 513–531.

[93] Gassert, R.; Metzger, J.-C.; Leuenberger, K.; Popp, W. L.; Tucker, M. R.; Vigaru, B.; Zimmermann, R.; Lambercy, O. Physical student–robot interaction with the ETHZ haptic paddle. *IEEE Trans. Educ.* **2013**, *56*, 9–17.

[94] Murphy, K.; Darrah, M. Haptics-Based Apps for Middle School Students with Visual Impairments. *IEEE Trans. Haptic* **2015**, *8*, 318–326.

[95] Reiner, M. Conceptual construction of fields through tactile interface. *Interact. Learn. Environ.* **1999**, *7*, 31–55.

[96] Zhang, J.; Johnson, K. A.; Malin, J. T.; Smith, J. W. Human-centered information visualization. International workshop on dynamic visualizations and learning, Tubingen, Germany. 2002.

[97] Davis, R. L.; Martinez, M. O.; Schneider, O.; MacLean, K. E.; Okamura, A. M.; Blikstein, P. The haptic bridge: Towards a theory for haptic-supported learning. Proc. 2017 Conf. Interact. Des. Child. 2017; pp 51–60.

[98] Rau, M. A. Conditions for the effectiveness of multiple visual representations in enhancing STEM learning. *Educ. Psychol. Rev.* **2017**, *29*, 717–761.

[99] Mayer, R. E. *The Cambridge Handbook of Multimedia Learning*; Cambridge University Press, 2005; pp 31–48.

[100] Gilbert, J. K.; Treagust, D. F. *Multiple representations in chemical education*; Springer, 2009; Vol. 4.

[101] Moreno, R.; Mayer, R. Interactive multimodal learning environments: Special issue on interactive learning environments: Contemporary issues and trends. *Educ. Psychol. Rev.* **2007**, *19*, 309–326.

[102] Ainsworth, S. DeFT: A conceptual framework for considering learning with multiple representations. *Learn. Instr.* **2006**, *16*, 183–198.

[103] Eberle, F. Studienerfolg von Absolventinnen und Absolventen der Ergänzungsprüfung "Passerelle" an den universitären Hochschulen: Studie im Auftrag der Schweizerischen Maturitätskommission. *Schriftenreihe SBFI* **2022**,

[104] Schwartz, D. L.; Martin, T. Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cogn. Instr.* **2004**, *22*, 129–184.

[105] Kapur, M. Examining productive failure, productive success, unproductive failure, and unproductive success in learning. *Educ. Psychol.* **2016**, *51*, 289–299.

[106] Glogger-Frey, I.; Fleischer, C.; Grüny, L.; Kappich, J.; Renkl, A. Inventing a solution and studying a worked solution prepare differently for learning from direct instruction. *Learn. Instr.* **2015**, *39*, 72–87.

[107] Lamnina, M.; Chase, C. C. Developing a thirst for knowledge: How uncertainty in the classroom influences curiosity, affect, learning, and transfer. *Contemp. Educ. Psychol.* **2019**, *59*, 101785.

[108] Loibl, K.; Rummel, N. Knowing what you don't know makes failure productive. *Learn. Instr.* **2014**, *34*, 74–85.

[109] Sinha, T.; Kapur, M.; West, R.; Catasta, M.; Hauswirth, M.; Trninic, D. Differential benefits of explicit failure-driven and success-driven scaffolding in problem-solving prior to instruction. *J. Educ. Psychol.* **2021**, *113*, 530.

[110] Sinha, T.; Kapur, M. Robust effects of the efficacy of explicit failure-driven scaffolding in problem-solving prior to instruction: A replication and extension. *Learn. Instr.* **2021**, *75*, 101488.

[111] Murayama, J.; Shimizu, H.; Nam, C. S.; Satoh, H.; Sato, M. An educational environment for chemical contents with haptic interaction. **2007**, 346–352.

[112] Keller, J. M. *Motivational design for learning and performance: The ARCS model approach*; Springer, 2010; Chapter The Arcs model of motivational design, pp 43–74.

[113] Lanrezac, A.; Frey, N.; Baaden, M. Wielding the power of interactive molecular simulations. *WIREs Comput. Mol. Sci.* **2022**, *12*, e1594.

[114] Valle, M. Visualization: A cognition amplifier. *Int. J. Quantum Chem.* **2013**, *113*, 2040–2052.

[115] Brooks, Jr, F. P. Impressions by a dinosaur – Summary of Faraday discussion 169: Molecular simulations and visualization. *Faraday Discuss.* **2014**, *169*, 521–527.

[116] Levinthal, C. Molecular model-building by computer. *Sci. Am.* **1966**, *214*, 42–52.

[117] Brooks, F. P.; Ouh-Young, M.; Batter, J. J.; Kilpatrick, P. J. Project GROPE Haptic displays for scientific visualization. *SIGGRAPH Comput. Graph.* **1990**, *24*, 177–185.

[118] Atkinson, W. D.; Bond, K. E.; Tribble, G. L.; Wilson, K. R. Computing with feeling. *Comput. Graphics* **1977**, *2*, 97–103.

[119] Křenek, A. Haptic rendering of complex force fields. Proc. Workshop Virt. Environ. 2003; p 231V239.

[120] Lee, Y.-G.; Lyons, K. W. Smoothing haptic interaction using molecular force calculations. *Comput.-Aided Des.* **2004**, *36*, 75–90.

[121] Daunay, B.; Régnier, S. Stable six degrees of freedom haptic feedback for flexible ligand-protein docking. *Comput.-Aided Des.* **2009**, *41*, 886–895.

[122] Bolopion, A.; Cagneau, B.; Redon, S.; Régnier, S. Variable gain haptic coupling for molecular simulation. IEEE World Haptic Conf. 2011; pp 469 – 474.

[123] Stocks, M. B.; Hayward, S.; Laycock, S. D. Interacting with the biomolecular solvent accessible surface via a haptic feedback device. *BMC Struct. Biol.* **2009**, *9*, 69.

[124] Ouh-young, M.; Pique, M.; Hughes, J.; Srinivasan, N.; Brooks, F. Using a manipulator for force display in molecular docking. IEEE Rob. Autom. 1988; pp 1824–1829.

[125] Levine, D.; Facello, M.; Hallstrom, P.; Reeder, G.; Walenz, B.; Stevens, F. Stalk: An interactive system for virtual molecular docking. *IEEE Comput. Sci. Eng.* **1997**, *4*, 55–65.

[126] Nagata, H.; Mizushima, H.; Tanaka, H. Concept and prototype of protein-ligand docking simulator with force feedback technology. *Bioinformatics* **2002**, *18*, 140–146.

[127] Wollacott, A. M.; Merz Jr, K. M. Haptic applications for molecular structure manipulation. *J. Mol. Graphics Modell.* **2007**, *25*, 801–805.

[128] Sourina, O.; Torres, J.; Wang, J. *Transactions on Edutainment II*; Springer, 2009; pp 105–118.

[129] Comai, S.; Mazza, D. A haptic-enhanced system for molecular sensing. Interact. 2009; pp 493–496.

[130] Surles, M. C.; Richardson, J. S.; Richardson, D. C.; Jr., F. P. B. Sculpting proteins interactively: Continual energy minimization embedded in a graphical modeling system. *Protein Sci.* **1994**, *3*, 198–210.

[131] Mark, W. R.; Randolph, S. C.; Finch, M.; Verth, J. M. V.; Taylor, R. M. Adding force feedback to graphics systems: Issues and solutions. Annu. Conf. Comput. Graphics Interact. Tech. 1996; p 447V452.

[132] Ruspini, D. C.; Kolarov, K.; Khatib, O. The haptic display of complex graphical environments. Annu. Conf. Comput. Graphics Interact. Tech. 1997; pp 345–352.

[133] Anthopoulos, A.; Pasqualetto, G.; Grimstead, I.; Brancale, A. Haptic-driven, interactive drug design: Implementing a GPU-based approach to evaluate the induced fit effect. *Faraday Discuss.* **2014**, *169*, 323–342.

[134] Iakovou, G.; Hayward, S.; Laycock, S. D. Virtual environment for studying the docking interactions of rigid biomolecules with haptics. *J. Chem. Inf. Model.* **2017**, *57*, 1142–1152.

[135] Jaillet, L.; Artemova, S.; Redon, S. IM-UFF: Extending the universal force field for interactive molecular modeling. *J. Mol. Graphics Modell.* **2017**, *77*, 350–362.

[136] Matthews, N.; Kitao, A.; Laycock, S.; Hayward, S. Haptic-assisted interactive molecular docking incorporating receptor flexibility. *J. Chem. Inf. Model.* **2019**, *59*, 2900–2912.

[137] Bosson, M.; Grudinin, S.; Redon, S. Block-adaptive quantum mechanics: An adaptive divide-and-conquer approach to interactive quantum chemistry. *J. Comput. Chem.* **2013**, *34*, 492–504.

[138] Dubois, M.-A.; Bouju, X.; Rochefort, A. Toward interactive scanning tunneling microscopy simulations of large-scale molecular systems in real time. *J. Appl. Phys.* **2018**, *124*, 044301.

[139] Haag, M. P.; Marti, K. H.; Reiher, M. Generation of potential energy surfaces in high dimensions and their haptic exploration. *ChemPhysChem* **2011**, *12*, 3204–3213.

[140] Haag, M. P.; Reiher, M. Studying chemical reactivity in a virtual environment. *Faraday Discuss.* **2014**, *169*, 89–118.

[141] Porezag, D.; Frauenheim, T.; Köhler, T.; Seifert, G.; Kaschner, R. Construction of tight-binding-like potentials on the basis of density-functional theory: Application to carbon. *Phys. Rev. B* **1995**, *51*, 12947.

[142] Seifert, G.; Porezag, D.; Frauenheim, T. Calculations of molecules, clusters, and solids with a simplified LCAO-DFT-LDA scheme. *Int. J. Quantum Chem.* **1996**, *58*, 185–192.

[143] Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B* **1998**, *58*, 7260.

[144] Gaus, M.; Cui, Q.; Elstner, M. DFTB3: Extension of the self-consistent-charge density-functional tight-binding method (SCC-DFTB). *J. Chem. Theory Comput.* **2011**, *7*, 931–948.

[145] Dewar, M. J.; Thiel, W. Ground states of molecules. 38. The MNDO method. Approximations and parameters. *J. Am. Chem. Soc.* **1977**, *99*, 4899–4907.

[146] Thiel, W.; Voityuk, A. A. Extension of the MNDO formalism to d orbitals: integral approximations and preliminary numerical results. *Theor. Chim. Acta* **1992**, *81*, 391–404.

[147] Dewar, M. J.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.

[148] Rocha, G. B.; Freire, R. O.; Simas, A. M.; Stewart, J. J. Rm1: A reparameterization of am1 for h, c, n, o, p, s, f, cl, br, and i. *J. Comput. Chem.* **2006**, *27*, 1101–1111.

[149] Stewart, J. J. P. Optimization of parameters for semiempirical methods I. Method. *J. Comput. Chem.* **1989**, *10*, 209–220.

[150] Stewart, J. J. P. Optimization of parameters for semiempirical methods II. Applications. *J. Comput. Chem.* **1989**, *10*, 221–264.

[151] Stewart, J. J. P. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *J. Mol. Model.* **2007**, *13*, 1173–1213.

[152] Husch, T.; Vaucher, A. C.; Reiher, M. Semiempirical molecular orbital models based on the neglect of diatomic differential overlap approximation. *Int. J. Quantum Chem.* **2018**, *118*, e25799.

[153] Vaucher, A. C. Interactive control of quantum mechanical calculations in chemistry. Ph.D. thesis, 2018.

[154] Behler, J. Perspective: Machine learning potentials for atomistic simulations. *J. Chem. Phys.* **2016**, *145*.

[155] Amabilino, S.; Bratholm, L. A.; Bennie, S. J.; Vaucher, A. C.; Reiher, M.; Glowacki, D. R. Training neural nets to learn reactive potential energy surfaces using interactive quantum chemistry in virtual reality. *J. Phys. Chem. A* **2019**, *123*, 4486–4499.

[156] Käser, S.; Vazquez-Salazar, L. I.; Meuwly, M.; Töpfer, K. Neural network potentials for chemistry: Concepts, applications and prospects. *Digital Discovery* **2023**, *2*, 28–58.

[157] Eckhoff, M.; Reiher, M. Lifelong machine learning potentials. *J. Chem. Theory Comput.* **2023**, *19*, 3509–3525.

[158] Fombona-Pascual, A.; Fombona, J.; Vicente, R. Augmented reality, A review of a way to represent and manipulate 3D chemical structures. *J. Chem. Inf. Model.* **2022**, *62*, 1863–1872.

[159] Mazzuco, A.; Krassmann, A. L.; Reategui, E.; Gomes, R. S. A systematic review of augmented reality in chemistry education. *Rev. Educ.* **2022**, *10*, e3325.

[160] Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849–857.

[161] Boobbyer, D. N.; Goodford, P. J.; McWhinnie, P. M.; Wade, R. C. New hydrogen-bond potentials for use in determining energetically favorable binding sites on molecules of known structure. *J. Med. Chem.* **1989**, *32*, 1083–1094.

[162] Reynolds, C. A.; Wade, R. C.; Goodford, P. J. Identifying targets for bioreductive agents: Using GRID to predict selective binding regions of proteins. *J. Mol. Graphics* **1989**, *7*, 103–108.

[163] Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

[164] Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098.

[165] Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *PCCP* **2005**, *7*, 3297–3305.

[166] Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

[167] Anderson, A. B. Electron density distribution functions and the ASED-MO theory. *Int. J. Quantum Chem.* **1994**, *49*, 581–589.

[168] Halgren, T. A. MMFF VI. MMFF94s option for energy minimization studies. *J. Comput. Chem.* **1999**, *20*, 720–729.

[169] Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; Van Gunsteren, W. F. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* **2011**, *40*, 843–856.

[170] Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* **2006**, *27*, 1787–1799.

[171] Tersoff, J.; Hamann, D. R. Theory of the scanning tunneling microscope. *Phys. Rev. B* **1985**, *31*, 805.

[172] Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.

[173] Ikeguchi, M.; Ueno, J.; Sato, M.; Kidera, A. Protein structural change upon ligand binding: Linear response theory. *Phys. Rev. Lett.* **2005**, *94*, 078102.

[174] O'Connor, M.; Deeks, H. M.; Dawn, E.; Metatla, O.; Roudaut, A.; Sutton, M.; Thomas, L. M.; Glowacki, B. R.; Sage, R.; Tew, P. Sampling molecular conformations and dynamics in a multiuser virtual reality framework. *Sci. Adv.* **2018**, *4*, eaat2731.

[175] Jones, M. G.; Andre, T.; Superfine, R.; Taylor, R. Learning at the nanoscale: The impact of students' use of remote microscopy on concepts of viruses, scale, and microscopy. *JRST* **2003**, *40*, 303–322.

[176] Hallman, G.; Paley, I.; Han, I.; Black, J. B. Possibilities of haptic feedback simulation for physics learning. Edmedia+ innovate learning. 2009; pp 3597–3602.

[177] Han, I.; Black, J. B. Incorporating haptic feedback in simulation for learning physics. *Comput. Educ.* **2011**, *57*, 2281–2290.

[178] Minogue, J.; Jones, G. Measuring the impact of haptic feedback using the SOLO taxonomy. *Int. J. Sci. Educ.* **2009**, *31*, 1359–1378.

[179] Jones, M. G.; Minogue, J.; Tretter, T. R.; Negishi, A.; Taylor, R. Haptic augmentation of science instruction: Does touch matter? *Sci. Educ.* **2006**, *90*, 111–123.

[180] Bivall, P.; Ainsworth, S.; Tibell, L. A. E. Do haptic representations help complex molecular learning? *Sci. Educ.* **2011**, *95*, 700–719.

[181] Schönborn, K. J.; Bivall, P.; Tibell, L. A. Exploring relationships between students' interaction and learning with a haptic virtual biomolecular model. *Comp. Educ.* **2011**, *57*, 2095–2105.

[182] Persson, P. B.; Cooper, M. D.; Tibell, L. A.; Ainsworth, S.; Ynnerman, A.; Jonsson, B.-H. Designing and evaluating a haptic system for biomolecular education. 2007 IEEE Virtual Real. Conf. 2007; pp 171–178.

[183] Wiebe, E. N.; Minogue, J.; Jones, M. G.; Cowley, J.; Krebs, D. Haptic feedback and students' learning about levers: Unraveling the effect of simulated touch. *Comp. Educ.* **2009**, *53*, 667–676.

[184] Ucar, E.; Ustunel, H.; Civelek, T.; Umut, I. Effects of using a force feedback haptic augmented simulation on the attitudes of the gifted students towards studying chemical bonds in virtual reality environment. *BIT* **2017**, *36*, 540–547.

[185] Edwards, B. I.; Bielawski, K. S.; Prada, R.; Cheok, A. D. Haptic virtual reality and immersive learning for enhanced organic chemistry instruction. *Virtual Reality* **2019**, *23*, 363–373.

[186] Mayer, R. E.; Moreno, R. Nine ways to reduce cognitive load in multimedia learning. *Educ. Psychol.* **2003**, *38*, 43–52.

[187] Crandell, O. M.; Lockhart, M. A.; Cooper, M. M. Arrows on the page are not a good gauge: Evidence for the importance of causal mechanistic explanations about nucleophilic substitution in organic chemistry. *J. Chem. Educ.* **2020**, *97*, 313–327.

[188] Anzovino, M. E.; Bretz, S. L. Organic chemistry students' ideas about nucleophiles and electrophiles: the role of charges and mechanisms. *CERP* **2015**, *16*, 797–810.

[189] Cooper, M. M.; Klymkowsky, M. W. The trouble with chemical energy: Why understanding bond energies requires an interdisciplinary systems approach. *LSE* **2013**, *12*, 306–312.

[190] Aakre, I. G.; Persson, J. R.; Lein, H. L.; Eggen, P.-O. First-year university students' perception of chemical bonding and bond energy. *NJSTEME* **2021**, *4*.

[191] Becker, N. M.; Cooper, M. M. College chemistry students' understanding of potential energy in the context of atomic-molecular interactions. *JRST* **2014**, *51*, 789–808.

[192] Kohn, K. P.; Underwood, S. M.; Cooper, M. M. Energy connections and misconnections across chemistry and biology. *LSE* **2018**, *17*, ar3.

[193] Lancor, R. Using metaphor theory to examine conceptions of energy in biology, chemistry, and physics. *Sci. Educ.* **2014**, *23*, 1245–1267.

[194] Park, M.; Liu, X. Assessing understanding of the energy concept in different science disciplines. *Sci. Educ.* **2016**, *100*, 483–516.

[195] Özmen, H. Some student misconceptions in chemistry: A literature review of chemical bonding. *J. Sci. Educ. Tech.* **2004**, *13*, 147–159.

[196] Prodjosantoso, A. K.; Hertina, A. M. The misconception diagnosis on ionic and covalent bonds concepts with three tier diagnostic test. *Int. J. Instr.* **2019**, *12*, 1477–1488.

[197] Taber, K. S. Building the structural concepts of chemistry: Some considerations from educational research. *CERP* **2001**, *2*, 123–158.

[198] Taber, K. S. College students' conceptions of chemical stability: The widespread adoption of a heuristic rule out of context and beyond its range of application. *Int. J. Sci. Educ,* **2009**, *31*, 1333–1358.

[199] Gilbert, J. K.; De Jong, O.; Justi, R.; Treagust, D. F.; Van Driel, J. H. *Chemical education: Towards research-based practice*; Springer Science & Business Media, 2003; Vol. 17.

[200] Boo, H. K. Students' understandings of chemical bonds and the energetics of chemical reactions. *JRST* **1998**, *35*, 569–581.

[201] Kelly, R. M.; Barrera, J. H.; Mohamed, S. C. An analysis of undergraduate general chemistry students' misconceptions of the submicroscopic level of precipitation reactions. *J. Chem. Educ.* **2010**, *87*, 113–118.

[202] Pérez García, M.; Luxford, C. J.; Windus, T. L.; Holme, T. A quantum chemistry concept inventory for physical chemistry classes. *J. Chem. Educ.* **2016**, *93*, 605–612.

[203] Galley, W. C. Exothermic bond breaking: A persistent misconception. *J. Chem. Educ.* **2004**, *81*, 523.

[204] Dreyfus, B. W.; Geller, B. D.; Gouvea, J.; Sawtelle, V.; Turpen, C.; Redish, E. F. Ontological metaphors for negative energy in an interdisciplinary context. *Phys. Rev. Spec. Top. Phys. Educ. Res.* **2014**, *10*, 020108.

[205] 3D Systems Touch Haptic Devices. https://de.3dsystems.com/haptics-devices/touch, 2022; [Online; accessed 04-October-2022].

[206] Itkowitz, B.; Handley, J.; Zhu, W. The OpenHaptics(TM) Toolkit: A Library for Adding 3D Touch (TM) Navigation and Haptics to Graphics Applications. 2005; pp 590–591.

[207] Peters, M.; Laeng, B.; Latham, K.; Jackson, M.; Zaiyouna, R.; Richardson, C. A redrawn Vandenberg and Kuse mental rotations test - Different versions and factors that affect performance. *Brain Cogn.* **1995**, *28*, 39–58.

[208] Ekstrom, R. B.; Dermen, D.; Harman, H. H. *Manual for Kit of Factor-Referenced Cognitive Tests*; Princeton, NJ: Educational Testing Service, 1976; Vol. 102.

[209] Brooke, J. SUS-A quick and dirty usability scale. *Usability evaluation in industry* **1996**, *189*, 4–7.

[210] Jennett, C.; Cox, A. L.; Cairns, P.; Dhoparee, S.; Epps, A.; Tijs, T.; Walton, A. Measuring and defining the experience of immersion in games. *Int. J. Hum. Comput.* **2008**, *66*, 641–661.

[211] Bangor, A.; Kortum, P. T.; Miller, J. T. An empirical evaluation of the system usability scale. *Int. J. Hum.-Comput. Interact.* **2008**, *24*, 574–594.

[212] Schmidt, M.; Kafka, J. X.; Kothgassner, O. D.; Hlavacs, H.; Beutl, L.; Felnhofer, A. Why does it always rain on me? Influence of gender and environmental factors on usability, technology related anxiety and immersion in virtual environments. ACE. 2013; pp 392–402.

[213] Rosenthal, R. Experimenter effects in behavioral research. **1976**,

[214] Braun, V.; Clarke, V. Using thematic analysis in psychology. *Qual. Res. Psychol.* **2006**, *3*, 77–101.

[215] Braun, V.; Clarke, V. *Thematic analysis: a practical guide*; SAGE, 2022.

[216] Cook, M. P. Visual representations in science education: The influence of prior knowledge and cognitive load theory on instructional design principles. *Sci. Educ.* **2006**, *90*, 1073–1091.

[217] Fiorella, L.; Mayer, R. E. Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. *The Cambridge handbook of multimedia learning* **2021**,

[218] Watson, D.; Clark, L. A.; Tellegen, A. Development and validation of brief measures of positive and negative affect: The PANAS scales. *J. Pers. Soc. Psychol.* **1988**, *54*, 1063–1070.

[219] Naylor, F. D. A state-trait curiosity inventory. *Aust. Psychol.* **1981**, *16*, 172–183.

[220] Glogger-Frey, I.; Gaus, K.; Renkl, A. Learning from direct instruction: Best prepared by several self-regulated or guided invention activities? *Learn. Instr.* **2017**, *51*, 26–35.

[221] Gonzalez-Franco, M.; Peck, T. C. Avatar embodiment. Towards a standardized questionnaire. *Front. Rob. AI* **2018**, *5*, 74.

[222] Hayes, A. F. *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach.*; 2013; Vol. 51.

[223] Muma, J. R. *Pragmatics of language*; Springer US, 1991; pp 229–247.

[224] Phillips, D. C.; Burbules, N. C. *Postpositivism and educational research*; Rowman & Littlefield, 2000.

[225] Mair, P.; Wilcox, R. Robust statistical methods in R using the WRS2 package. **2020**, *52*, 464–488.

[226] Rau, M. A. Comparing multiple theories about learning with physical and virtual representations: Conflicting or complementary effects? *Educ. Psychol. Rev.* **2020**, *32*, 297–325.

[227] Kalyuga, S.; Sweller, J. *The Cambridge Handbook of Multimedia Learning*; Cambridge University Press, 2014; pp 247–262.

[228] Mautone, P. D.; Mayer, R. E. Signaling as a cognitive guide in multimedia learning. *J. Educ. Psychol.* **2001**, *93*, 377–389.

[229] Scherr, R. E.; Close, H. G.; McKagan, S. B.; Vokos, S. Representing energy. I. Representing a substance ontology for energy. *Phys. Rev. Phys. Educ. Res.* **2012**, *8*, 020114.

[230] Nardo, A.; Trninic, D. Learning from Negative Experience: A Philosophical Exploration of 'Productive Failure'. **2020**,

[231] Mulford, D. R.; Robinson, W. R. An inventory for alternate conceptions among first-semester general chemistry students. *J. Chem. Educ.* **2002**, *79*, 739.

[232] Dempster, A. P.; Laird, N. M.; Rubin, D. B. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc., B* **1977**, *39*, 1–22.

[233] Schwarz, G. Estimating the dimension of a model. *Ann. Stat* **1978**, 461–464.

[234] Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716–723.

[235] McCutcheon, A. L. *Latent class analysis*; Sage, 1987.

[236] Linzer, D. A.; Lewis, J. B. poLCA: An R Package for Polytomous Variable Latent Class Analysis. *J. Stat. Softw.* **2011**, *42*, 1–29.

[237] Neese, F. The ORCA program system. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 73–78.

[238] Neese, F. Software update: the ORCA program system, version 4.0. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2018**, *8*, e1327.

[239] Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminformatics* **2012**, *4*, 1–17.

[240] Yuen, K. K. The two-sample trimmed t for unequal population variances. *Biometrika* **1974**, *61*, 165–170.

[241] Sharma, R.; Melcer, E. F.; Kao, D. Exploring relevance, meaningfulness, and perceived learning in entertainment games. 2022.

[242] Hendry, R. F. Two conceptions of the chemical bond. *Philos. Sci.* **2008**, *75*, 909–920.

[243] Magnasco, V. A model for the chemical bond. *J. Chem. Educ.* **2004**, *81*, 427.

[244] DeKock, R. L. The chemical bond. *J. Chem. Educ.* **1987**, *64*, 934.

[245] Kronik, L.; Levy Nahum, T.; Mamlok-Naaman, R.; Hofstein, A. A new "bottom-up" framework for teaching chemical bonding. *J. Chem. Educ.* **2008**, *85*, 1680.

[246] Stevens, S. Y.; Delgado, C.; Krajcik, J. S. Developing a hypothetical multi-dimensional learning progression for the nature of matter. *JRST* **2009**, *47*, 687–715.

[247] Stein, C. J.; Reiher, M. Automated selection of active orbital spaces. *J. Chem. Theory Comput.* **2016**, *12*, 1760–1771.

[248] Stein, C. J.; von Burg, V.; Reiher, M. The delicate balance of static and dynamic electron correlation. *J. Chem. Theory Comput.* **2016**, *12*, 3764–3773.

[249] Stein, C. J.; Reiher, M. Measuring multi-configurational character by orbital entanglement. *Mol. Phys.* **2017**, *115*, 2110–2119.

[250] Stein, C. J.; Reiher, M. Automated identification of relevant frontier orbitals for chemical compounds and processes. *Chimia* **2017**, *71*, 170.

[251] Stein, C. J.; Reiher, M. autoCAS: A program for fully automated multiconfigurational calculations. *J. Comput. Chem.* **2019**, *40*, 2216–2226.

[252] White, S. R. Density matrix formulation for quantum renormalization groups. *Phys. Rev. Lett.* **1992**, *69*, 2863–2866.

[253] White, S. R. Density-matrix algorithms for quantum renormalization groups. *Phys. Rev. B* **1993**, *48*, 10345–10356.

[254] Chan, G. K.-L.; Dorando, J. J.; Ghosh, D.; Hachmann, J.; Neuscamman, E.; Wang, H.; Yanai, T. *Frontiers in quantum systems in chemistry and physics*; Springer-Verlag, 2008; pp 49–65.

[255] Chan, G. K. L.; Zgid, D. The Density Matrix Renormalization Group in Quantum Chemistry. *Annu. Rep. Comput. Chem.* **2009**, *5*, 149–162.

[256] Marti, K. H.; Reiher, M. The Density Matrix Renormalization Group algorithm in quantum chemistry. *Z. Phys. Chem.* **2010**, *224*, 583–599.

[257] Schollwöck, U. The density-matrix renormalization group in the age of matrix product states. *Annu. Phys.* **2011**, *326*, 96–192.

[258] Chan, G. K.-L.; Sharma, S. The Density Matrix Renormalization Group in quantum chemistry. *Annu. Rev. Phys. Chem.* **2011**, *62*, 465–481.

[259] Wouters, S.; Van Neck, D. The Density Matrix Renormalization Group for ab initio quantum chemistry. *Eur. Phys. J. D* **2013**, *31*, 272.

[260] Kurashige, Y. Multireference electron correlation methods with density matrix renormalisation group reference functions. *Mol. Phys.* **2014**, *112*, 1485–1494.

[261] Olivares-Amaya, R.; Hu, W.; Nakatani, N.; Sharma, S.; Yang, J.; Chan, G. K.-L. The ab-initio Density Matrix Renormalization Group in practice. *J. Chem. Phys.* **2015**, *142*, 34102.

[262] Szalay, S.; Pfeffer, M.; Murg, V.; Barcza, G.; Verstraete, F.; Schneider, R.; Legeza, Ö. Tensor product methods and entanglement optimization for ab initio quantum chemistry. *Int. J. Quantum Chem.* **2015**, *115*, 1342–1391.

[263] Yanai, T.; Kurashige, Y.; Mizukami, W.; Chalupský, J.; Lan, T. N.; Saitow, M. Density matrix renormalization group for ab initio calculations and associated dynamic correlation methods: A review of theory and applications. *Int. J. Quantum Chem.* **2015**, *115*, 283–299.

[264] Baiardi, A.; Reiher, M. The Density Matrix Renormalization Group in chemistry and molecular physics: Recent developments and new challenges. *J. Chem. Phys.* **2020**, *152*, 040903.

[265] Brockmole, J. R.; Davoli, C. C.; Abrams, R. A.; Witt, J. K. The world within reach: Effects of hand posture and tool use on visual cognition. *Curr. Dir. Psychol. Sci.* **2013**, *22*, 38–44.

[266] O'Connor, M. B.; Bennie, S. J.; Deeks, H. M.; Jamieson-Binnie, A.; Jones, A. J.; Shannon, R. J.; Walters, R.; Mitchell, T. J.; Mulholland, A. J.; Glowacki, D. R. Interactive molecular dynamics in virtual reality from quantum chemistry to drug binding: An open-source multi-person framework. *J. Chem. Phys.* **2019**, *150*, 220901.

[267] Chatain, J.; Sisserman, D. M.; Reichardt, L.; Fayolle, V.; Kapur, M.; Sumner, R. W.; Zünd, F.; Bermano, A. H. Digiglo: Exploring the palm as an input

and display mechanism through digital gloves. Proc. Annu. Symp. Comput. Hum. Interact. Play. 2020; pp 374–385.

[268] Jorion, N.; Self, B.; James, K.; Schroeder, L.; DiBello, L.; Pellegrino, J. Classical test theory analysis of the dynamics concept inventory. Proc. 2013 Am. Soc. Eng. Educ. 2013.

[269] Briggs, S. R.; Cheek, J. M. The role of factor analysis in the development and evaluation of personality scales. *J. Pers.* **1986**, *54*, 106–148.

Acknowledgments

I had the privilege to complete this work in two insanely supportive research groups and this privilege comes with a long list of people that need to be thanked.

First and foremost, I thank my two **supervisors**, *Prof. Dr. Markus Reiher* and *Prof. Dr. Manu Kapur* for trusting me with this project and supporting me throughout the years. Your guidance provided the scaffold needed in this interdisciplinary endeavor while giving me the space to explore and learn.

My learning process was further facilitated by the **senior members** of the research groups. I am forever grateful to *Venera Gashaj* who continues to be a mentor and a dear friend even after her departure from the group. Your often invisible work is indispensable. Additionally, I want to express my gratitude to *Thomas Weymuth* and *Jan Unsleber* for spending countless hours in GUI meetings and code reviews. The art of doing learning science research was taught to me by *Tanmay Sinha*. Thank you for the effort that you put into teaching students from domains across the board. *Aline Nardo*, thank you for introducing me to a more philosophical perspective on learning and education. Your inputs played a pivotal role in my personal and professional growth. Lastly, I am grateful for the calm and kind presence of *Alberto Baiardi* and *Dragan Trninic*, the interactions with whom always left me grounded and content.

Especially at the beginning of my doctorate, I relied strongly on the infinite wisdom of the – at the time – **senior PhD students**. I wholeheartedly thank *Francesco Bosia* and *Christoph Brunken* for their endless patience in introducing me to the SCINE framework. *Tobias Halbherr* set the bar extremely high when it comes to learning study designs. Your critical feedback was always greatly appreciated. My very first contact with the Reiher group was with *Vera von Burg*. From that first coffee with me as a master student to the end I always appreciated your active, compassionate listening. I am also deeply grateful to *Stephanie Grimmel* for encouraging me every step of the way and for providing me with relaxant tea in critical times. Finally, I thank *Jan-Grimo Sobez* for introducing me to Yotam Ottolenghi. The culunary education that you provided before leaving the group is deeply valued.

Especially after spending too many hours in home office, I appreciated the lively discussions, mutual support, active feedback culture and coffee/ice cream breaks with my

fellow doctoral students. *Julia Chatain*, you have been an amazing friend, mentor, collaborator and most importantly boxing partner. *Nina Glaser*, thank you for keeping my phone out of my bedroom and also more generally for the never-ending reservoir of genuinely good advice. *Katja Csizi* has deeply impressed me with her debating skills but also her resilient and calm way to navigate stressful situations (finishing a thesis in record time and navigating the Hardbrücke intersection, to name two). Thank you for answering my endless stream of questions especially towards the end. To *Paul Türtscher*, thank you for offering pragmatic solutions whenever I unnecessarily complicated my life and *Miguel Steiner*, thank you for the amazing gorgonzola pasta and the countless improvements made on my code (both to be valued equally). *Max Mörchen*'s research feels like art to me and while I appreciate and admire CC diagrams thanks to you, I will try to eventually also understand them. I highly appreciated *Stefan Gugler*'s uncompromising questions. Discussions with you always left me deeply confused in the exactly right way. *Samuel Tobler*, thank you for the constant flow of new research and presentation tools. We will always be one step behind you. *Vera Baumgartner*, thank you for including me in your friend group. I hope to see you on June 14th (year unspecified). *Nadja Beeler*, my health significantly improved as I climbed more stairs, thanks to you. Thank you *Christian Fässler* for welcoming me in your home and introducing me to your lovely family. With *Cléa Formaz* I share the love for writing. Your feedback was always highly appreciated. It is thanks to *Robin Feldmann* that I will forever question my conception of the molecular structure. Your nontraditional and I dare to say philosophical approaches to standing concepts helped me to understand deeper but also discover my own knowledge gaps.

I also thank *Sara May*, *Siara Isaac*, *Helena Kovacs* and *Leon Calabrese* for supporting this project from **behind the scenes**. Thank you Sara, for the occasional reminder email and several coffee machine saves. Thank you Siara and Helena for introducing me to qualitative methods and hosting me at EPFL. Thank you Leon, for transcribing my first interview studies. Interviewing is indeed a skill that needs to be learned and I am aware that my first attempts were not always easy to work with.

I am also grateful to the **newer members** of the groups. To *Moritz Bensberg*, for always making sure that we are not trapped in the stream of students at lunch time. To *Veronika Klasovita*, for making sure the temperature in the office is never too cold and to *Marco Eckhoff*, for his leveled and diplomatic nature, it brings balance to the group. *Fan Wang*, it was amazing to witness how you developed unique ideas during your master thesis that will now become the basis for an interesting PhD project. And finally, to *Mihael Erakovic*, *Enric Pérez* and *Kalman Szenes*, I wish you a fruitful, engaging and productive time in the Reiher group.

Finally, my deepest appreciation goes to my **family** for supporting me since the beginning. To my siblings, *Madeleine*, *Dominik* and *Bettina*, thank you for inspiring me with your diverse journeys. I am excited to see what unexpected junctions you will take, what mountains you will climb and most importantly, what breaks you will take. To my dear parents, *Susann* and *Thomas*, thank you for always having my back and providing any book that I have ever needed. And to *Julia*, thank you for your care during times of overwhelming to-do lists. I will never take it for granted.

Publications

The following publications are included in parts of this thesis:

- Müller, C. H.; Reiher, M.; Kapur, M. Embodied activities as preparation for learning quantum chemistry: A mixed-method study. *J. Comput. Assist. Learn.* **In press**.

Additionally, other parts of this thesis may be published at a later time.

During my PhD I contributed to the following software:

- Bensberg, M.; Brandino, G. P.; Can, Y.; Del, M.; Grimmel, S. A.; Mesiti, M.; Müller, C. H.; Steiner, M.; Türtscher, P. L.; Unsleber, J. P.; Weberndorfer, M.; Weymuth, T.; Reiher, M. qcscine/heron. *Zenodo*, **2022**, https://doi.org/10.5281/zenodo.7038388.
- Baiardi, A.; Bensberg, M.; Bosia, F.; Brunken, C.; Csizi, K.-S.; Feldmann, R.; Glaser, N.; Grimmel, S. A.; Gugler, S.; Haag, M. P.; Heuer, M. A.; Müller, C. H.; Polonius, S.; Simm, G. N.; Sobez, J.-G.; Steiner, M.; Türtscher, P. L.; Unsleber, J. P.; Vaucher, A. C.; Weymuth, T.; Reiher, M. qcscine/utilities. *Zenodo*, **2023**, https://doi.org/10.5281/zenodo.7928050.
- Bosia, F.; Husch, T.; Müller, C. H.; Polonius, S.; Sobez, J.-G.; Steiner, M.; Unsleber, J. P.; Vaucher, A. C.; Weymuth, T.; Reiher, M. qcscine/sparrow. *Zenodo*, **2023**, https://doi.org/10.5281/zenodo.7928079.

The following publication covers the research I did before I started my PhD at ETH Zurich:

- Knochenmuss, R.; Maity, S.; Balmer, F.; Müller, C.; Leutwyler, S.; Intermolecular dissociation energies of 1-naphthol · n-alkane complexes. *J. Chem. Phys.*, **2018**, *149*, 034306.

Charlotte H. Müller

Personal details

Date of birth:	April 19, 1995
Place of birth:	Bern, Switzerland
Nationality:	Swiss
Orcid:	0000-0002-6640-5065

Education

02.2020 – 09.2023	Doctoral Studies in Theoretical Chemistry and Learning Sciences, ETH Zurich, Switzerland *Thesis:* Embodied Quantum Chemistry Learning from Haptic Feedback *Supervisors:* Prof. Dr. Markus Reiher & Prof. Dr. Manu Kapur
08.2019 – 01.2020	CAS Scientific Journalism, MAZ – The Swiss School for Journalism
09.2017 – 06.2019	MSc Computational Science, University of Zurich, Switzerland *Thesis:* Localization of Kohn-Sham Orbitals via QR Decomposition *Supervisor:* Prof. Dr. Jürg Hutter
09.2013 – 06.2017	BSc Chemistry and Molecular Science, University of Bern, Switzerland *Thesis:* UV-Spectroscopy of Jet-Cooled 1-Methyl-5,6-Trimethylene Cytosine and its H_2O Complex *Supervisor:* Prof. Dr. Samuel Leutwyler

Employment

11.2023 – Today	*Postdoctoral Researcher*, ETH Zurich, Switzerland *Supervisor:* Prof. Dr. Martina Rau
02.2020 – 09.2023	*Doctoral Researcher*, ETH Zurich, Switzerland *Supervisor:* Prof. Dr. Markus Reiher & Prof. Dr. Manu Kapur
06.2019 – 08.2019	*IAESTE Internship*, Koç Üniversitesi *Supervisor:* Prof. Dr. Özgür Esat Müstecaplıoğlu
08.2016 – 02.2017	*Research Assistant*, University of Bern *Supervisor:* Prof. Dr. Samuel Leutwyler

Teaching Activities

02.2020 – 06.2023	*Praktikum Physikalische & Analytische Chemie,* *Versuch: Quantenchemische Studien*, ETH Zurich
09.2020 – 12.2022	*Informatik I*, ETH Zurich
01.2019 – 12.2021	ABQ Schulprojekt
09.2016 – 12.2016	*Physical Chemistry III*, University of Bern

Approved Projects

02.2020 – 06.2023	*Project:* Facilitating Learning of Energy & Forces in Chemistry via Incorporation of the Haptic Modality *Approved by:* ETH Ethics Commission

Awards

09.2021	Runner-up Poster Award SCS Fall Meeting 2021
10.2018 – 04.2019	INSPIRE Potentials – MARVEL Master's Fellowship

Scientific Talks

2022	Physical Chemistry Colloquium, ETH Zurich *Title:* Real–time Haptic Quantum Chemistry for Chemistry Education
2022	Future Learning Initiative Colloquium, ETH Zurich *Title:* Learning Chemical Concepts from Haptic Feedback – A Qualitative Study
2021	Behavioural Studies Colloquium, ETH Zurich *Title:* Haptic Quantum Chemistry: How Feeling Forces can Facilitate Learning about Reactivity
2020	Future Learning Initiative Collquium, ETH Zurich *Title:* Haptic Quantum Chemistry: How Feeling Forces can Facilitate Learning about Reactivity

Conference Contributions

2023	EARLI – Poster *Title:* Concreteness in Quantum Chemistry
2023	Berliner Methodentreffen Qualitative Forschung – Poster *Title:* Die vielfältigen Konzeptionen der chemischen Bindung bei Studierenden: Eine Fallstudie im Kontext der einführenden Quantenchemie-Vorlesung
2022	SCS Fall Meeting – Poster *Title:* Bachelor Students' Understanding of Basic Quantum Chemical Concepts
2022	EARLI SIG 6 & 7 Conference – Technology Demonstration *Title:* SCINE Interactive: A Quantum–based Virtual Learning Environment for Chemistry with Haptic Feedback
2022	ISLS Annual Meeting – Doctoral Consortium *Title:* Facilitating Learning of Quantum Chemical Concepts through Grounding in Sensory Experience
2021	SCS Fall Meeting – Poster *Title:* Real-Time Haptic Chemistry: Dive Hands–First into the Molecular World
2021	JURE – Poster *Title:* Access the Molecular World through Haptic Quantum Chemistry