# Performance and Participation in Scientific Experimentation

## An Exploration of Human Thinking as Mental Action, from Scientific Experimentation to Systematic Thought Experiments

*Renatus Ziegler*

*Summary*

Experimentation is at the core of the natural sciences. Experiments are done for different reasons. They differ substantially from observations, where naturally occurring processes are taken note of or measured. Experiments instantiate the specific conditions that are necessary for isolating an experimental setting from its environment and manipulating the course of events in some way. This paper focuses on the role of human thinking as a mental action in the process of experimentation. In many cases, human action, in particular mental action, is taken for granted during experimentation, and the analysis tends to focus on questions of causality, the type of instruments and materials used, etc. This paper proposes to broaden the focus in the analysis of scientific experimentation, by taking mental actions into consideration: first by considering thought experiments, then pure thought experiments, and eventually what we might call (pure) experiments in thinking. Thinking is something that all human beings are capable of and, therefore, one can explore it systematically oneself. This paper outlines what the opportunities for such first-person experimentation are and how they can contribute to the enquiry of scientific experimentation.

*Zusammenfassung*

Experimentieren gehört wesentlich zu den Naturwissenschaften. Experimente dienen verschiedenen Zielen. Sie unterscheiden sich deutlich von blossen Beobachtungen, bei denen natürlich auftretende Prozesse registriert und gemessen werden. In Experimenten werden spezifische notwendige/hinreichende Bedingungen festgehalten, um die Experimentieranordnung von der Umgebung zu isolieren und um darüber hinaus den Verlauf der Ereignisse in dieser oder jener Weise zu beeinflussen. Der Fokus im vorliegenden Artikel liegt auf der Rolle des tätigen Denkens im Prozess des Experimentierens. Üblicherweise wird das tätige Denken als gegeben hingenommen und die Analyse des Experimentierens konzentriert sich auf Fragen der Kausalität, der verwendeten Instrumente und Materialien usw. Es wird

vorgeschlagen, den Fokus der Analyse des Experimentierens zu erweitern, insbesondere durch den Einbezug der performativen Qualität des tätigen Denkens, indem zusätzlich Gedankenexperimente, dann Experimente des reinen Denkens einbezogen werden, und zuletzt dasjenige, was hier reine Gedankenexperimente genannt wird. Denken ist eine Fähigkeit, die allen Menschen zur Verfügung steht, und folglich sind Menschen in der Lage, dies bei sich selbst systematisch zu untersuchen. Es wird dargestellt, welche Möglichkeiten und Herausforderungen solche Experimente aus der Erste-Person-Perspektive bieten und was sie zur Untersuchung wissenschaftlichen Experimentierens beitragen können.

## *1. Introduction*

Experimentation belongs to the core of many fields in natural science (*Gooding* 1990, p. xi; *Gooding et al.* 1989, p. xiii). Experiments are done for different reasons: to evaluate or refute theories, to explore fields of research, to find regularities, to build concepts, to design and test instruments, etc. They differ substantially from observations, where naturally occurring processes are accompanied by taking measurements, making notes, and building theories or, at least, concepts around them. The use of sophisticated technology, such as is done in astronomy, does not automatically turn an observation into an intervention (*Carrier* 1998, p. 176). Without doubt, an observation can be (and in many cases actually is) an intervention, but as such it still plays no role in causing, bringing about, or directing the observed process. In contrast, the set-up of any experiment instantiates the specific initial conditions that are necessary and sufficient for the experiment to work, including isolating the experimental setting from its environment and manipulating the course of events in one way or another.

This paper focuses on a somewhat unusual and neglected issue: the role of human thinking as a mental action that is an integral, performative part of the process of experimentation. What does it mean for mental activity to perform an experiment in contrast to just observing something?

Conventionally, human action, in particular mental action, is rarely discussed explicitly or in any detail within the philosophy of scientific experimentation, with the notable early exception of *Dingler* (1928, part II, Ch. I, §3). Sometimes it is even excluded in the philosophical analysis of experiments (*Hon* 1998, pp. 228, 233). Human action, particularly mental action, is often tacitly taken for granted as an inherent part of experimentation and therefore the analysis focuses, in most cases, on the issue of causality, on the type of instruments and materials used (*Hacking* 1992; *Heidelberger* 1998), on laboratory practices (*Steinle* 2005), on the exploration of new fields, on the problem of theory-ladeness (*Heidelberger* 2003), etc. See, e.g., the overviews (*Gooding et al.* 1989; *Heidelberger*

2009; *Heidelberger* & *Steinle* 1998; *Radder* 2003; 2009). According to this understanding, what we might consider to be a categorical difference between scientific experimentation and the observation of natural occurrences and chains of events, can get lost. Of particular interest in this context is the replication of an experiment, which is, on the one hand, dependent on human action (particularly mental or thinking action) and, on the other hand, specifically tries to overcome the human factor.

To broaden the focus in the analysis of scientific experimentation, this paper proposes to first consider thought experiments, then pure thought experiments and eventually what we might call (pure) experiments in thinking. The combined exploration of these subjects helps in evaluating the common features, similarities, and differences between them; in reflecting on them alongside each other, they show their specific features more clearly.

Taking the human factor into consideration in scientific experimentation can be accomplished in different ways. One way looks at the consequences for, and interrelations with, technology and the implications for the environment and the living world (Lebenswelt), etc. The focus here, however, is on human beings as agents, not subjects, of scientific experimentation.

This is where the phenomenology of thinking and the subject of mental actions come into play: implicitly in the descriptions of various experimental settings and explicitly at the end of this paper. Thinking is something one can do and hence can explore systematically oneself, namely through experiments in thinking which will be discussed later. It has recently been shown (*Breyer* & *Gutland* 2016a; *Pitt* 2004), that the phenomenology of thinking is very much dependent on detailed and clear-cut first-person accounts. It will be demonstrated here how first-person accounts of such thinking processes, which are present in all scientific experimentation, may be explored in more depth.

## *2. Scientific experimentation*

### *2.1. Phases of scientific experimentation*

To begin, the focus will be on purely scientific experimentation, that is, on experiments that are done in isolation from the natural and technical environment. This isolation – the experimental and explicit control of external factors and the exclusion of chance events – is an important ingredient of any such experimentation. In fact, trying to construct a closed system is one of the main features of what constitutes scientific experimentation.

There are three main phases in the execution of a scientific experiment (*Lange* 2003, pp. 121f.), that, however, might be subdivided further (*Hacking* 1992; *Janich* 1998, pp. 102–107): (1) the set-up and artificial

variation of initial conditions, including control of the environment; (2) the triggering or initiation of the process or situation that one wants to focus on; (3) the occurrent response, process or situation that at the very least can be tracked and measured, and in some cases observed or perceived directly (see Table 1).

| | **Design phase** | **Set-up of initial conditions and control of environment (isolation)** | **Triggering of events / processes** | **Ensuing events / processes** | **Events / processes measurable and / or observable** | **Reference to empirical facts** | **Visualization, mental imagery** |
|---|---|---|---|---|---|---|---|
| **Natural scientific experiment** | Performative action (thought experiment) | Performative action | Performative action | Occurring events / processes | Yes | Explicit | Yes, in the design and interpretative phase |
| **Computer simulation** | Performative action (thought experiment) | Performative action | Performative action | Occurring events / processes | Yes, in the output | Implicit (implementation of laws of nature) | Yes, in the design and interpretative phase |
| **Thought experiment** | – | Performative action | Performative action | Performative action | Not during action: after completion of action | Implicit (implementation of laws of nature) | Yes, in all phases |
| **Pure thought experiment** | – | Performative action | Performative action | Performative action | Not during action: after completion of action | No | No |
| **Pure experiment in thinking** | – | Performative action | Performative action | Performative action | Not during action: after completion of action | No | No |

*Table 1:* Performative action in the phases of different types of experiments

Depending on the specific circumstances, these phases may overlap. If the necessary initial conditions and their artificial variations are sufficient (as in experiments within classical optics: refraction, reflection, diffraction etc.), then (1) immediately implies (2). If the control of the environment has to be maintained until all intended measurements or observations are completed, then (1) might last until (3) is finished.

There is a fourth part to an experiment, namely the interpretation of the results and the contextualization into current scientific concepts and theories of the set-up, its variations and the outcome or response. This has been the subject of wide discussion concerning the nature of experiments and their relation to theoretical elements. See, e.g.: explorative vs. confirmatory experiments (*Steinle* 1997; 1998), experiments depending on symbolic constructions (*Duhem* 1998, Ch. 8, 10), and basic experiments with minimal and/or elementary theoretical frameworks (*Duhem* 1998; *Hacking* 1983; 1992). The distinction between empirical laws and full-

blown theories is covered by several authors (*Heidelberger* 1998; *Steinle* 2005, pp. 309–316). The debate on this point is not relevant here since it does not affect the discussion about the performative parts that constitute any type of experiment in contrast to mere observation of natural processes.

There is also a preliminary part of scientific experimentation that will be discussed later, namely the design and goal setting phase, which includes questions about the relations of the variables one is planning to vary and the corresponding outcome.

In the view presented here, scientific experimentation has a productive or performative part, and a receptive or observational part (*Heidelberger* 2009, pp. 165f., 175f.; *Hon* 1998, pp. 239–242; *Janich* 1998; 2016; *Radder* 1998, pp. 396–398; *Steinle* 2005, pp. 301–305, 312; *Tetens* 1987, pp. 2–4, 12, 17–19). The experimenters produce or set up and vary the initial conditions, construct an isolated system by controlling the environment, and then watch and measure the outcome/response. In short, experiments are controlled manipulation of nature (*Buchwald* 1998, pp. 378–382) in order to enable nature to show its own regularities.

Georg Henrik von Wright writes:

> *"I would maintain that we cannot understand causation, nor the distinction between nomic connections and accidental uniformities of nature, without resorting to ideas of doing things and intentionally interfering with the course of nature. […] [it] is convenient to distinguish between* doing *things and* bringing about *things and therefore also between ability to do and ability to bring about."*
>
> (*Wright* 1971, pp. 65f.)

This is exactly what one has to have in mind in differentiating between the above-mentioned phases (1) and (2), on the one hand, and phase (3) on the other. David Gooding emphasizes the "active intervention", the "importance of human agency" and the "procedural turn", where one looks at experiments as organized into sequences of acts and operations (*Gooding* 1990, pp. 8–14; 1992b, pp. 45, 47, 52–57).

We need not go as far as *Hacking* (1983, pp. 220–232) and postulate that experimenters produce the outcome of their experiments in a strict sense – rather we might think that experimenters bring about the experimental results. The experimenter is responsible for the set-up and control, whereas the result and/or the process that it leads to is an event that is encountered, that happens to the experimenter, that is given to him relative to the performative action (*Gooding* 1992a, pp. 65–69), and can only be acknowledged by a receptive mode of consciousness.

If there were no performative parts in experimentation, the difference between an experiment and the mere observation of a natural process or a natural chain of events, including the use of scientific measurement

instruments, would vanish and with it the control of the initial conditions and the environment.

This further implies that the causes of a performative action cannot be natural processes alone – purely natural chains of events, with no clear-cut human initiation – because then, essentially, they would have to be observable processes in all details, which they are not (*Glymour* 2004; *Janich* 1998, pp. 102–107; *Ziegler* 2003). The experimenter is *doing* something in setting up and isolating the experiment from the environment and not simply observing a process that sets up the experiment and isolates it using given natural processes that may be accepted or rejected.

Primas puts it as follows:

> *"The idea of measurement [in experimentation]* presupposes *that we are able to discriminate between the observing and the observed system.* Our ability to describe the world cannot go farther than our ability to isolate objects. *The separation of the universe into two parts, a part* which sees *and a part* which is seen*, is a conceptual necessity […]."*
>
> (*Primas* 1991, p. 334)

If this were not the case, no experiment could occur, no isolation from an environment could be constructed, no control could be performed, and there would be no reliable result. We would not be able to differentiate with certainty which factors are relevant for the outcome and which factors are just confounding the result.

Hence, two very different kinds of causes are present in any scientific experiment: on the one hand, the performative action of the experimenter, including thinking action (setting up initial conditions, triggering the process) and, on the other hand, the natural causes that are controlled along with the ones that govern the observable/measurable outcome.

## *2.2. Replication*

This paper aims to corroborate this last claim of the dual nature of scientific experimentation (performative action including thinking action vs. observation/measurement). Replication is one of the cornerstones of the experimental sciences: an experiment that cannot be replicated is not an experiment in the strict sense. Without doubt, there are natural processes that cannot be replicated in experimental settings, such as the movement of the planets or other bodies in the solar system. However, gravitational and other forces that are present in these situations can be studied in laboratories on earth by replicable experiments (otherwise no science of the solar system would be possible). There may also be technological and financial constraints in replicating a very complex experiment that involves hundreds of co-workers, technicians, engineers, IT-specialists, etc.

Concerning the specific initial conditions set up by the experimenter, the claim for replication goes directly against the specific nature of natural chains of events: due to the multitude of factors determining them, natural occurrences never take up the same initial conditions themselves, hence they never strictly replicate themselves.

To be sure, this article does not want to imply that within laboratory science, that is, the world of experiments, there are no natural chains of events – there are. We initiate and enable them, otherwise we could not explore and experience nature. However, natural chains of events need to be prevented from crossing the boundary from outside into the laboratory and, in particular, into the experiments. This is what "isolation" means: something that disturbs, or rather disrupts the universal course of all chains of events by setting up initial conditions that are not part of this chain of events.

Scientists will typically not wait until an extremely rare event of natural replication occurs. Since strict replication presupposes isolation (control of the environment), something natural processes do not do, such a replicated event occurring naturally is highly improbable if not outright impossible. But we do not need to speculate about such natural replications: scientists *do* replicate experiments, emphasize their importance, and are, as a rule, quite good at it.

But this comes at a cost, or rather, in this view, with an added feature: the replication of experiments is based on performative actions of the experimenter (*Steinle* 2005, pp. 307–309) in setting up the initial conditions (and their artificial variation) and controlling the environment, that is, the isolation – and as such disrupt or go against the universal course of natural or occurrent chains of events.

### *2.3. Failure and error in experiments*

Failures happen – there is no need to explain this in detail (for more on this, however, see below). On the other hand, natural processes never fail, they just happen, according to their own natural laws. In addition, an observation (without any kind of theoretical interpretation) never fails as such, given that all observational devices (senses, instruments) work well. The observation simply yields the results it shows – it is neither wrong nor right. Observing something wrongly (excluding interpretational errors) can only mean that one does not observe it at all or observes it only partially, or without properly functioning instruments/senses. However, this depends entirely on the performative action of the observer, which can fail or not – and here we are again.

Hence, failures occur (in most cases) since, as experimenters, we fail by imperfectly controlling the environment, by using the wrong kinds of measurement devices, by using broken instruments, by taking insufficient

recordings, by misinterpreting the set-up and/or the results, etc. Because of this, most authors analyzing failures and errors in experimentation seek to find criteria and methodological safeguards to avoid them (*Allchin* 2001; *Hon* 1998), or they evaluate the epistemic roles of errors in particular (*Schickore* 2005). This is as performative as it gets: natural processes, in particular physical processes, never try to avoid something, safeguard themselves against failure or the like; they never isolate themselves from the environment on their own (whatever that might mean).

To put it quite succinctly: failures or errors occur if, and only if, we fail in some way to exercise performative action in setting up, controlling and triggering experimental conditions and if we give way to a natural or psychological process (including relying naively on conventions, prejudices, associations, intuitions, norms, paradigms etc.) and thus disturb the strict and unbiased performance.

We do not, and cannot, fail or err by intention. To fail/err is one thing, it happens as long as one is not aware of it. Everything else is deception or outright fraud (a very performative undertaking indeed).

To *discover* an error/failure, to realize that something has gone amiss, is quite another matter: we are confronted with challenging facts that cannot be ignored; the performative action has run up against something that needs to be taken into account. This throws us back on ourselves, to the analysis of the performative action, which shows, eventually, that we – not nature – have failed. In this sense, the performative action is not, at least not in all of its aspects, part of occurrent natural processes.

Before an experimenter sets up his or her experiment, they need to think about what they want to achieve, what kind of initial conditions they need, what control mechanisms they are supposed to work out, what triggers the process they want to examine, and what kind of measuring devices to apply. Given all this, we try to run through the experiment mentally and anticipate the processes and possible results. In other words, we execute a *thought experiment*, in which all components are present as particular thoughts, concepts, mental images etc. and are composed in such a way that, in principle, the relevant experiment could be done. Hence, in its design phase, any real scientific experiment is based on a thought experiment, and all initial conditions, etc., need to be set up first in the realm of thought before they can be executed accurately in reality.

## *3. Thought experiments*

### *3.1. The nature of thought experiments*

"Thought experiments are conducted in mental laboratories but they do not thereby cease to be experiments." (*Gooding* 1992c, p. 281; see also *Buzzoni* 2008; *Hopp* 2014). Thought experiments are scientific

experiments in the sense that, in most cases, they have a target thesis and set-up, and are characterized by artificial variation of clear-cut initial conditions (the scenario, based on background knowledge). They bring about an experimental process (the application of the scenario to some specific situation) and a result (the outcome of this application). They are performative; they are processes "to be worked through." "Personal participation is essential: it is what makes a thought experiment an *experiment* rather than another form of argumentation." (*Gooding* 1992c, pp. 285, 281; see also *Gooding* 1990, pp. 203–205, 215).

A discussion of additional processes for assessing thought experiments, namely interpretation, contextualization, and formulation of conclusions, is beyond the scope of this paper. Additionally, this paper does not discuss their reliability, the process of knowledge acquisition, or the sources of the respective knowledge (*Wiltsche* 2018). It is important to note that thought experiments are based on a broad field of embodied experience and knowledge of scientific theories, and one needs to feel at home in such experiences/knowledge (*Fehige* & *Wiltsche* 2013; *Gooding* 1992c, p. 285). "Finally, like real experiments in their demonstrative or textbook form, thought experiments always work. Their narratives have the inexorable character of geometrical demonstrations." (*Gooding* 1992b, p. 72).

The focus in this paper is on the performative aspect of thought experiments. Usually, overviews on thought experiments do not mention their performative aspects as a research subject (*Stuart et al.* 2018), or if they do, they do so only in passing (*Sorensen* 1992, pp. 219, 223, 241).

### *3.2. The Twin-Earth example*

Here we will examine the Twin-Earth thought experiment (*Putnam* 1973; 1975) as a standard example. The set-up includes the assumption of the existence of a Twin-Earth, a planet far out in the cosmos with identical features to Earth and an environment, biosphere, and human beings that are no different than those on Earth in their physical aspects, right down to the molecular level. The inhabitants of Twin-Earth even speak the same language(s), have the same proper names, etc. The only difference is that they use the word "water" to refer to a substance with a complicated structure, called XYZ, that has the same *observable* properties as what Earthlings call "water", i.e. $H_2O$. However, $H_2O$ does not exist on Twin-Earth, and XYZ does not exist on Earth. It is further assumed that the inhabitants of these two planets do not know about each other, and therefore do not know that what Earthlings call "water" and what Twin-Earthlings call "water" refer to two different substances, namely $H_2O$ and XYZ respectively. So much for the scenario.

This is now applied to the question of the source of the meaning of words. People on both planets learn to use the word "water" through

having the same experiences with the same substance and perhaps through learning about its chemical structure. Consistent with the scenario, the overall behavior in the interactions with water on both planets should be the same. This is the main "observation": there is no need to know about the specific structural differences between $H_2O$ and XYZ in order to understand the behavior of human beings on Earth and on Twin-Earth. But are they referring to the same object by using the word "water"? It is not necessary to go into the details of the ongoing debate, since the main interest in this paper is not to evaluate how it is possible (if at all) to draw conclusions for the meaning of "meaning" from such thought experiments, but rather to ask: what kind of performative action is involved in such experiments?

### *3.3. Performative aspects in thought experiments*

The set-up of thought experiments is highly performative. To perform a thought experiment "means to successfully 'live through'" this experiment (*Wiltsche* 2018, pp. 357, 359f.): it is neither drawn from already observable facts nor are its specific details taken from worked-out theories alone – it is purely thought-out, imagined, and visualized (*Gooding* 1992c; *Nersessian* 2018). The set-up of such an experiment is not an observable natural process but something done by the experimenter: it has to be brought up and maintained in a directive, executive sense, not by just triggering something that occurs without further intervention or keeps going without intentional force (see Table 1).

As a rule, the set-up of thought experiments is speculative at least, plausible at best, and more or less probable. In this sense, it relies on present (and past) knowledge and inferences from that. However, the performative composition of this knowledge goes beyond what is already known.

Equally, what happens when applying the thought-out scenario to specific circumstances or questions has to be done or executed – it has to be performed. This is not a mere event or occurrence happening separately from acting human beings.

Hence, contrary to natural scientific experiments, *all* phases of a thought experiment have performative aspects. However, they also rely on common sense, naturalness, plausibility and occurrent intuitions about the known world, and hence have a mixed character: the performative actions, mainly the compositional part (the scenario), are based on receptive processes and intuitions, including prior or presupposed knowledge, helpful associations, mostly from the natural sciences.

Contrary to scientific experiments, where one relies on direct evidence and natural causality, thought experiments rely on plausibility or indirect (thought-out, hypothetical, not present) evidence. E.g., the scenario of Twin-Earth assumes the same biology for all humans on both planets. This might

be highly improbable, given the different nature of water on each planet, and taking into account the role of water in biology in general and the fact that the adult human organism contains around 70 % water. In addition, the identity of human beings on Earth and Twin-Earth presupposes that with an identical molecular configuration, their psychic and behavioral features would be the same, which might be controversial as well.

### *3.4. Replication and failure in thought experiments*

In accordance with natural scientific experiments, thought experiments can be replicated (given the same intuitions) and can fail – two features which corroborate their performative character. Excluding outright logical errors, failure or error in thought experiments is mainly due to non-generalizable intuitions, implausible arrangements of known facts, etc. The detection of an error and its correction is not as clear-cut as in scientific experimentation, but is performative, to say the least. Errors happen by themselves – but by becoming aware of them, we are confronted with faulty arguments or different levels of intuitions, which we may correct or not. Not all authors agree on the same intuitions, even if they are based on natural science (see the discussion of the Twin-Earth example above), and errors do not repair themselves without individual human action.

| | **Certainty** | **Initial awareness of experimental events / processes** | **Initial awareness of set-up** | **Focus, subject matter** |
|---|---|---|---|---|
| **Natural scientific experiment** | Evidential, causal | Receptive | Performative | Set-up, triggering, manipulated objects, isolation |
| **Computer simulation** | Plausibility (depending on known laws of nature) | Receptive | Performative | Premises, inference rules |
| **Thought experiment** | Plausibility (depending on known laws of nature) | Performative | Performative | Scenario, visualizations, common knowledge |
| **Pure thought experiment** | Consistency (logical, mathematical) | Performative | Performative | Conceptual insight, conceptual relations, inferences |
| **Pure experiment in thinking** | Experiential within performance | Performative | Performative | Conceptual process, actions, events within actions |

*Table 2:* Characteristic qualitative features of different types of experiments

Thought experiments are peculiar in that they rely on strict argumentations but highly debatable intuitions. In scientific experiments, we do not need to discuss plausibility or the reliability of intuitions theoretically, that is, without any reference to the experiments – apart from basic ingredients such as consistency and falsifiability – we just execute or replicate the experiment and that settles any doubts (see Table 2). In this respect, scientific experiments, as a rule, are *pure*: their content (set-up, trigger, result) does not rely on debatable intuitions, non-measurable facts, non-occurrent processes, or events/processes that are not accounted for. What might be debatable in this case are the sweeping theoretical frameworks on which one performs such experiments. However, this article is not the place to discuss this, since the focus lies on specific experimental facts which are more or less independent from such frameworks: theories come and go, basic experimental facts remain.

On the other hand, the performance that brings about experimental content in thought experiments is pure enablement, in the sense that it, ideally, does not interfere with the content. In other words, such experiments are replicable and as such can be inter-subjectively validated.

### *3.5. Design and performative action in thought experiments*

Before closing this section on thought experiments, it is necessary to point out a characteristic quality of thought experiments in comparison to natural scientific experiments. In the latter, there is the necessity of a design phase *prior* to the execution of the experiment, otherwise we would not know what to do in general nor in detail. By contrast, the design phase in thought experiments, particularly in the mind of the designer (who is the experimenter in the view presented here), coincides with the beginning of the execution of the experiment. Certainly, there might be preliminary deliberations leading towards the thought experiment, and we may need to run through them several times until they fit the purpose; but to be aware of what we want to think about means to already be performing the thought experiment. In other words, we cannot design a consistent thinking process without thinking it. Hence, there is no clear-cut separation between the design phase and the execution phase for a thought experiment (Table 1).

However, in contrast to the design executed by oneself, if we take up the design of a thought experiment given verbally by someone else, we might consider, e.g., the instruction "think that $2 \times 2 = 5$", without actually thinking or reflecting on it. Mere verbal intake of an instruction does not fit the case that is analyzed here: setting up and reflecting on the design is an integral part of any true experimental action and its replication, in the sense intended here.

For an assessment of thought experiments as compared to computer simulation experiments, see *Lenhard* (2018) and Tables 1 and 2.

At this point, the issue that brought this line of argumentation to thought experiments, namely the question of the nature of performative action, cannot be carried much further. We have seen that thought experiments rely on performative actions in *all* their phases, and in some sense are mixed with occurrent thoughts and intuitions.

Since both happen in the thinking mind, the quality of thoughts and the quality of performative actions in thought experiments are more closely linked to each other than in scientific experiments, where thinking does not play an explicit role in the content or in the subject matter of the ensuing process that is observed/measured (although it does in the interpretation of its result).

Is it possible to examine the performative action that presents itself in scientific experiments and in thought experiments as such? Can we examine performative actions not only as instruments enabling something else but enabling themselves? Can performative action be the target of enquiry of experiments done by performative actions? Some preliminary proposals regarding these questions will be given in the next sections.

Thought experiments were discussed in order to come closer to what performative action is and how it may be experienced. It is one step towards cleansing our actions of non-performative elements, leading up to pure thought experiments in order to get to the core of human mental action as something essentially apart from natural (causal) chains of events.

## *4. Pure thought experiments*

A further step is needed to explore the nature of performative actions, particularly thinking actions, within scientific experimentation: performative action within thinking must be made the direct object of inquiry. For this, isolating the experimental setting from all factors outside of performative action, and outside of its reach and interventional capacity, is suggested. Therefore, we are now looking for a *pure* thought experiment, in other words, one in which there is no reference – explicit, implicit, or inferred – to empirical evidence/facts or empirical knowledge. Let's start out with an example and discuss it in reference to its performative aspects.

### *4.1. Example of a pure thought experiment*

The proposal is to explore the rational numbers by first reflecting on a fundamental theorem of arithmetic, namely, that every integer can be factored into prime numbers (i.e., numbers that can only be divided by themselves or 1). This implies that every rational number $p'/q'$ with $q' \neq 0$ can be represented by a ratio $p/q$ where $p$ and $q$ ($\neq 0$) have no common factor, i.e. all common factors in both of the prime factorizations of $p'$ and $q'$ have been cancelled out.

To work with this as a pure thought experiment, it is important to immerse ourselves afresh, with a new mindset, every time we run through the example. We have to become aware of these arithmetical relations by going through them again and again, understanding them "right now" and not relying, as much as possible, on previous knowledge. This means that we do not just collect what we already know but that we think these relations actively each time, as if it were the first time.

What does this mean? To think something anew as if it was the first time – not just remembering past processes or executing it in an automatic manner – is just what we would expect in true replication of a natural scientific experiment. A replication must be conducted in such a way that, at least in principle, it is still open to what the outcome might be. In replication, we have to be very careful not to repeat past mistakes, misconceptions, observational errors, measurement errors, handling failures, etc. This is only possible, if we set up the replication experiment as if it were for the first time. Only then is there some guarantee that the replication is not just a repetition of past biases, insufficient isolation, etc.; only then could a replication possibly be the source of a new discovery or a new effect. Any time such an experiment is performed, it should have the potential to disrupt the path of science. Thus, a replication of an experiment should never be a routine but something that engages our full attention and that can surprise us again and again (even if it is the same as before).

Let us take up the example again, with the above in mind. After immersing ourselves in some elementary arithmetic of rational numbers, we might encounter the question of the nature of the square root of a prime number. Consider the square root of 2, which we may write as *sqrt*(2): is it a rational number? If we intentionally let go of any prior knowledge, we do not know the answer at first. Since we live very comfortable with rational numbers and can express any measurement with all the accuracy we need (using maybe a long but finite number of digits after the decimal point), there is no evidence to the contrary. However, we can try out the hypothesis: we can suppose that *sqrt*(2) *is* a rational number, say $p/q$ with no common factor, and probe the assumption.

According to the assumption, then, $2 = p^2/q^2$, hence $2q^2 = p^2$. This implies that $p^2$ is an even number (a number with a factor 2 in it), since the square of any odd number, $2m+1$, is always odd, that is, contains no factor 2: $(2m+1)^2 = 4m^2 + 4m + 1$. For this reason, let $p$ be an even number, namely $p = 2k$, $k \neq 0$; and given this, then from $2q^2 = p^2 = (2k)^2 = 4k^2$ we get $q^2 = 2k^2$, and hence $q^2$ has to be even too, using the same argument as before. This contradicts the unproblematic assumption that $p$ and $q$ have no common factors, as now they have one, namely 2. This should have come as a shock, since nothing was wrong in the calculation (please check!). However, there was an assumption here, namely that *sqrt*(2) is a rational number. The only possible conclusion is that *sqrt*(2) is *not* a

rational number, and so it belongs to a totally new category of numbers, the irrational numbers, which encompass even more peculiar numbers, namely transcendental numbers, like $e$ or $\pi$ that cannot be represented by roots of any kind.

It is important to note that this argument only works as a proof for a performing agent who goes through it, lives through it, and makes a mental effort (*Antonini* & *Mariotti* 2008, p. 402). The human agent has to take the assumptions seriously without anticipating the result, work straight through to the contradiction, take this logical fact as something that is not in his or her capacity to alter or shape, and thus draw the unavoidable conclusions (*Dutilh Novaes* 2016, p. 2618). We only *understand* the proof if we do it actively ourself, not just take it to be something we remember or know already or have learned at some point by whatever means. In the latter case, we may have understood the proof at some point *earlier* in life but this has no relevance for the *present* insight or understanding, since we *understand* something, namely conceptual relations, not because we remember them, but because we *think* them right now.

Another example that has been worked out in some detail elsewhere (*Ziegler* & *Weger* 2019) explores the purely conceptual relations of points, segments, lines, and circles. What does it mean to build the concept of a circle? Our interest here is not a specific circle (token) but the conceptual structure (type) of it. If we start with a point $M$ and a plane $\mathcal{E}$, the point $M$ either coincides with $\mathcal{E}$ or not. When we take all points in the plane $\mathcal{E}$ that have the same Euclidean distance $d$ from the point $M$ (somewhere in space), we arrive at what we call a circle with a middle point $M$ and a radius length smaller or equal to $d$, independent of the specific position of the plane $\mathcal{E}$, the position of point $M$ in space, and the length of distance $d$. We need no mental pictures or images to "see" this, it simply follows from what the relation between points, distances, and planes is. Or, as is often the case, we may argue the other way round: *if* there are images from unknown sources, then they can only be determined, analyzed, corroborated or refuted by these conceptual relations; conversely, the conceptual relations cannot be analyzed, etc., by them. Variation of the conceptual relations leads to different results. E.g., if we look for all points in space with the same distance from $M$, then a sphere results and we discover that the circle worked out above is the intersection of the given plane $\mathcal{E}$ with this sphere.

### *4.2. Why are these mathematical examples thought experiments?*

The examples do not fit the usual classification of thought experiments: they do not involve explicit reference to empirical knowledge or direct bodily experiences, and they do not reference visualizations or the like, which might occur in these cases but are not relevant for the argumentation. (For a different view on mathematical thought experiments involving

visualizations, see *Starikova* & *Giaquinto* 2018.) So are we not looking at an argumentation rather than an experimentation?

We need to differentiate between arguments as an assortment of thoughts with a given argumentative structure – say some propositions linked together by a deductive process – and the performance of such an argumentation. The latter may be based in total on this structure. However, it makes a huge difference if I just take note or know that such an argument exists in written form, such as a computer program, or if I actually perform it myself and thereby understand its content, structure and implications (see the examples above). In this sense, performative argumentation has an experimental quality: it includes a set-up (initial propositions, premises, and assumptions), implementation (running through the argument) and a result. All three phases are performative – even the result does not appear out of the blue, it has to be worked out. As *Buzzoni* (2011, pp. 75f.) puts it: "the anticipation in thought of the solution of a problem in pure mathematics amounts to its actual solution". In addition, there is a difference between conceptual insight, which needs performative action, and judgements that also rely on accessible knowledge, conventions or already known paths of argumentation (*Gutland* 2021).

Furthermore, pure thought experiments can fail in the same way as thought experiments can fail and one is able to detect such failures, correct them and run through the corrected pure thought experiment again.

### *4.3. Pure thought experiments as mental performances*

This paper set out to evaluate the performative aspect of scientific experimentation and has considered conventional thought experiments and pure thought experiments. When is a pure thought experiment a scientific experiment? Concerning the performative character of thought experiments, David Gooding remarks (already cited above), "Personal participation is essential: it is what makes a thought experiment an experiment rather than another form of argumentation." (*Gooding* 1992c, p. 281). Let us therefore summarize some characteristic features of pure thought experiments which are implied in the given examples – examples which should not be taken in their specific setting (as tokens) but as instantiations of the structure and the type of pure thought experiments which are considered in this paper. *Ziegler* & *Weger* (2018; 2019) have called this type of thinking "focused productive conceptual thinking" or "active thinking" for short.

(1) First, there are clear-cut initial conditions: we set out with specific concepts and goals (proof, exploration). (2) Second, through the execution or implementation of the goals that were determined, we delve deeper into the realm of conceptual relations of which the initial concepts are a part. (3) Third, we arrive at a result where either the proof is finished or the exploration stops and takes account of what has been established.

Here we see the similarities between natural scientific experimentation and pure thought experiments: an initial phase, an implementation phase, and a result phase (see Table 1 and 2).

However, there are also significant differences. There is no design phase in pure thought experiments separable from the initial phase, as is the case with conventional thought experiments: in thinking about the design, the pure thought experiment has already started or at least partially started if we use verbal or written instructions at the outset. Additionally, the implementation phase is not qualitatively separate from the initial phase: we just carry it out, persevering in what we set out to do, starting with the initial conditions (*Buzzoni* 2011, pp. 77–79). In accordance with thought experiments, this implementation process is not observable in the sense that we could sit back and wait, measure, and inspect what happens. It is not an occurrence of unknown origin but something of its own making (to be taken up later): the performance of human agents.

The most important feature of the type of pure thought experiments presented here, setting them distinctly apart from natural scientific experiments and thought experiments, is their *purity*. This refers to their structural independence from direct or indirect reference to sense-perceptions or the like. The concepts, conceptual relations or judgements we are considering (mainly from mathematics) do not rely on any specific sensual instantiations or mental pictures of them. By fixing the initial conditions, specific images (of lines, circles, numbers, etc.) may come to mind, but if we want to think in pure concepts, these images need to be abandoned quickly: they have no determinate and necessary influence whatsoever on further thought, which is only governed by the conceptual content and not mental images. Independence from direct or indirect sense-perceptions is further corroborated by the fact that this type of pure thought experiment needs no reference to any kind of plausibility: it does not depend on past, present, or future results of the natural sciences as do thought experiments, as a rule. All that counts are *consistency* and *coherence*, where consistency means the absence of contradictions (in the logical sense) and coherence means the eventual compatibility with results of all other pure thought experiments from logic and/or mathematics.

What perfect isolation, i.e. control of external factors and the environment, is for natural scientific experimentation, isolation from direct or indirect influences of sensual perceptions, moods, feelings, volitions etc. is for pure thought experiments. This might be difficult to achieve. It can be disturbed by diversions or distractions from external sources (human interaction, noises, etc.) or internal sources (brain-waves, volitions, pain, etc.), and known diversions from the path of thinking by mind-wandering (*Weger et al.* 2018). However, through training and perseverance, strong attentional engagement, etc., this can be overcome, in the same way that the execution of (nearly) perfectly isolated experiments can be achieved by trained, experienced and clear, goal-oriented researchers.

### *4.4. Error in pure thought experiments*

Error or failure in pure thought experiments happens when there is imperfect isolation and uninvited "guests" show up, i.e., prejudices, existing knowledge, mental representations, imaginations, volitions, etc., or when moods, feelings, etc., play a role in determining the initial conditions or the outcome. This means that the performative action that is aimed at pure concepts and conceptual relations is disturbed by occurrent events that are not part of the *performative* thought process.

To detect an error requires experiencing something within the active thinking process that needs to be taken into account: it cannot be made to vanish and cannot be changed or altered – it is simply there and has to be taken up.

## *5. Objections and validation*

Skepticism to this approach mainly concerns the concept and experiential evaluation of pure thought experiments (see the foregoing chapter) and experiments in pure thinking (see the following chapter). Both go beyond the usual phenomenological approach that is grounded in sense perception or the like (speech, mental images, etc.)

Does one have access to pure thoughts and to one's own mental action? Several alternative explanations for the phenomenology of thinking have been proposed. One common proposal, e.g., reduces the phenomenology of thinking to a bundle of experiences, bodily or otherwise, tied to sensations (*Tye* & *Wright* 2011). The authors write: "From a phenomenological perspective, thinking a thought is much like running a sentence through one's head and/or (in some cases) having a mental image in mind together with (in some cases) an emotional/bodily response and a feeling of effort if the thought is complex or difficult to grasp." (p. 329) Certainly, the issue is not whether such events happen during thinking or influence it (they do, see above); rather, the question is whether they *determine* the process of thinking and its content, and whether they sufficiently explain what happens while one thinks.

In the most conspicuous of these alternative explanations, namely that language, or sentences, are the most prominent means of consciously experiencing thought, several authors have argued that experiencing language should not be confused with *understanding* the conceptual content expressed by the words and sentences. E.g.: "Now, when we focus our attention on something by means of inner speech, we become conscious, not of our own inner speech – which is already a conscious phenomenon – but of the *content* of our inner speech. […] Inner speech, by driving our attention to contents, makes these contents conscious." (*Jorba* & *Vicente* 2014, p. 95).

Others contend that thinking action is nothing but a construction of the

brain, determined by brain processes. Thinking action is surely associated with a functioning brain, but a functioning brain is not sufficient to produce thinking. There are many periods in the life of a functioning brain where no active thinking occurs. On the other hand, if the brain is damaged, in most cases thinking action also does not take place. From such observations it follows that the brain is a necessary but not sufficient condition for thinking actions: it enables thinking but does not cause or determine it. This is corroborated by the fact that in conceptual thinking no explicit use is made of neurological facts or processes. To illustrate: in order to explain the conceptual content of, say, the arithmetical expression "9 = 2 + 7" (or the proof of the irrationality of *sqrt*(2) as above), one does not need to take into account neurological laws or specific results from neurological experiments – up until now, no one has shown that this is necessary in order to understand these concepts. If it were otherwise, any systematic, analytic, coherent and conceptually consistent philosophical or mathematical paper or book would need a basic chapter on neurology in order to be conceptually complete, logical and understandable.

As to the issue of validation, to begin with, the first-person approach taken in this paper is not a major issue since all experiences, observations, decisions, inferences, reflections etc. are, in the end (and in the beginning), facts of first-person experience, i.e. experiences within the mind. They can be compared with the results of other researchers, they can be controlled by inner and outer experimentation and by variation of experimental parameters, etc. It is through repeated research activity, in particular, that there is a chance to overcome unexpected and involuntary biases from expectations, prejudices and the like.

We could see the introspective first-person perspective as a method to broaden the consideration of approaches to thinking action. It adds an inner or directly experiential and more immediate view to what is well known from an outside or third-person perspective. As soon as the focus is on processes rather than results, on unfolding rather than fixed states, it seems appropriate to look closely at what happens within the mind where these activities originate, or at least are processed. Actually, conducting introspective exploration argues pragmatically for the value of direct introspective experience and research.

Instead of providing a general theoretical defense of the first-person approach, its benefit, value, and validity, some relevant critical features are explicated; for the details of this debate in more general terms see the reviews (*Chudnoff* 2015, pp. 21–43; *Petitmengin* & *Bitpol* 2009; *Spener* 2011; *Weger* & *Wagemann* 2015a; 2015b).

Throughout the entire paper, methodological considerations play a prominent role and are applied repeatedly; they were suggested by *Breyer* & *Gutland* (2016a). To begin with, we need to differentiate between (a) living through or experiencing something in a pre-reflective mode; then

(b) noticing it or becoming aware of it; and finally (c) communicating it (presenting it, writing about it). E.g., at first, we might be tempted to think that we realize only by reflection that something has been experienced. However, this turns out to not be true: there is a definite quality experienced in thinking conceptual relations, as has been shown above and elsewhere (*Ziegler* & *Weger* 2019). With respect to communication, we need to further clarify what happened during active thinking and translate this to written or spoken language to express it to the scientific community. Secondly, the distinction between generalizable and non-generalizable facts is important in this paper. We need to become aware of what makes the thinking experience, or the noticing of it, vary from person to person and what influences personal thinking: the environment (i.e. influences from other people), education, cognitive abilities, training, habituation, previous knowledge etc. In short, the main methodological question to pursue is: What are the relevant features of these experiments in thinking actions taken as types rather than token?

## *6. Experiments in Thinking*

### *6.1. What are experiments in thinking and why does one need them?*

This paper analyzes scientific experimentation and thought experiments mainly from the perspective of performative action – from a viewpoint that involves participation or engagement in executing such experiments. First, we have to look at what has been performed, what has been manipulated by the body or the thinking, i.e. at what these experiments are about – the subject matter or content. (Please note: with experiments in thinking, I am not referring to the recent excursions into experimental philosophy which, apart from being controversial, have a different focus: *Knobe et al.* 2012; *Knobe* & *Nichols* 2017; *Ludwig* 2007; 2018; *Nado* 2016; *Sosa* 2007; *Stich* & *Tobia* 2018.)

With *experiments in thinking* a radical shift of perspective occurs from "what are these experiments about?" to "how are these experiments performed?" including the mental set-up of the whole arrangement. This means focusing on the process, on the action as it is performed by human agents. This perspective is not new – researchers have already worked out certain aspects of it, e.g., the different performative stages pointed to in this paper: the set-up of an experiment, its execution, and the gathering of the results/measurements. In the different kinds of experiments discussed here, not all of these steps are performative (see Table 1 and 2); however, the set-up in all cases is not just a static occurrence that can be observed without participation but is initiated by the experimenter. It cannot be reduced to a causal process in the same sense as the causal processes explored in a scientific experiment. The cause (if one wants to call it that), or better, the

agent cause for the set-up is in the human experimenter, the human agent, not in a natural process into which this agent is embedded or embodied. Otherwise, there would be no experiment but only the occurrence of a natural chain of events which human agents are part of. However, experiments and their replication do not occur in nature: one does not, will not, and cannot wait for them to happen – they are performed, they are artifacts produced by human agents, otherwise, they would not exist. This makes them dependent on human agents with respect to their preliminary preparation and their appearance or their bringing about, but not with respect to their final essence or worked out subject matter (what they are about, their results, etc.).

In order to explore experimental set-ups more deeply, in particular the processes involved with this performative action, it was necessary to advance to experiments that rely even more on human action (in its broadest sense) than scientific experimentation, namely thought experiments and in particular pure thought experiments.

That is what this paper set out to do: produce and explore experimentation in order to evaluate the role of human agents and, in particular, thinking as a mental action. The more we can strip the performance of non-performative elements, namely from naturally occurring chains of events, the closer we come to the agentive action that is within all these performances. As was said, this allows us to explore an important aspect of experimentation, namely the role of thinking as a mental action, and thus gives us a more complete picture of the experiential psychology of experimentation.

Thus, the explorative path towards this goal, presented here, has led to pure thought experiments in which the entire experiment is a series of performances: nothing happens without involvement, no occurrence just drops in (besides disturbances, distractions, diversions, etc.), it all happens from the doing of the experimenter. But how does one know this? How could this be explored in more detail?

We are confronted with a hard problem, namely, how can we focus on the performance and at the same time be engaged in it? If we take an observation as something that is experienced without active involvement, that is, as something that is a receptive occurrence, then such performances cannot be observed. Well then, how can we explore them if there is nothing to observe? It has been worked out elsewhere that we can help ourself by "looking back" at what we have done and reflecting on it, *after* finishing the performance. Done systematically, this can lead to reliable results (*Ziegler & Weger* 2019). However, this seems to only be a preliminary way of exploring pure thought experiments.

Is there a way out of this post-hoc approach? A pure thinking process is not an action in the sense of setting some goal and then executing it, since the goal-setting process already involves thinking. And thinking cannot be observed, only done, otherwise one would have to split the personality into

an observer and a performer which is impossible (*Brentano* 1973, p. 181; *Comte* 1880, p. 25; *Fichte* 1997, p. 212; *Steiner* 1918, Ch. III).

The analysis of this situation will be presented in the next section but cannot do justice at this time to all the questions and issues concerning the subject matter of mental action, particularly mental performance within pure experiments in thinking. It was the primary purpose of this paper to show how deeply agentive performance pervades all types of scientific experimentation: without human agents there would be no experiments. And this pervasion, in its ultimate form, occurs within pure experiments in thinking: there is nothing non-agentive or non-active in this process. That is why this type of thinking must be the benchmark of agentive performances or activities against which all other types of experiments have to be measured.

### *6.2. Proposals from phenomenology using first-person approaches*

One needs to explore the experiential quality of these thinking processes as well as the experiential quality of the agency that initiates and executes them. What better place to explore how experiments in the natural sciences are initiated than to explore the realms where the thinking process that designs and initiates them has its experiential base and its source of agency, namely within pure experiments in thinking?

From the experiences with thinking processes, and in particular taking into account the research on scientific experimentation and thought experiments above, further explorative pathways are suggested that may help to tackle mental performances. This paper proposes to evaluate mental action and mental performance as it occurs in scientific experiments, thought experiments, and, in particular, pure thought experiments using the phenomenological method of first-person research (*Anderson* 2016; 2018; *Ziegler* & *Weger* 2018; 2019).

There are very few phenomenological accounts of thought experiments (*Fehige* & *Wiltsche* 2013; *Froese* & *Gallagher* 2010; *Hopp* 2014; *Wiltsche* 2018), and none of them takes up the subject of how the performative aspects – which are only mentioned explicitly by Walter Hopp and Harald Wiltsche – relate to mental activity or thinking activity, i.e., mental action.

In the following, some hints are given as to what may be promising approaches for tackling the experiential nature of mental action within the thinking process (see also *Steiner* 1918, Ch. III, VIII, IX; *Anderson* 2018, Ch. 7.) These approaches will not be defended or expanded here: further research has been done elsewhere (*Ziegler* & *Weger* 2023).

(1) As a first step, one needs to take into account the dynamic quality of the thinking process, the engagement or participation; for this one needs a first-person account since this cannot be inspected from the outside or

from a third-person perspective (*Ludwig* 2007). This includes the fact that such thinking processes are temporally extended, i.e. they take time or evolve in time: they unfold in time (*Ziegler* & *Weger* 2019, § 5.4).

(2) Second, since thinking actively is a multifaceted experience, first-person awareness has to be extended to the fringes of consciousness (*Mangan* 2001; *Petitmengin* & *Bitpol* 2009) or to the boundaries of focused awareness; several kinds of self-consciousness have to be taken into account (*Gallagher* & *Zahavi* 2013, pp. 52–58), particularly of the pre-reflexive or pre-predicative kind (*Ziegler* & *Weger* 2019, § 6.2). This includes the capacity to enhance conscious awareness in thinking in general (*Montague* 2016).

(3) Third, concerning action, the most appropriate and promising approach for evaluating thinking processes is the goal-oriented view of mental action (*Buckareff* 2005; *Mele* 1997; 2002). Specific thinking tasks are performed while one keeps executing the thinking activity in general. In this view, a thinking process is a complex undertaking with different phases and encompasses various tasks; these tasks and phases are focused or concentrated on a general theme or final purpose which we want to accomplish. With such a general goal in mind, we may be trying out different sub-approaches which might help us finding our way towards the main result – hence we need a teleological theory and an appeal for trying (*Proust* 2001).

(4) Fourth, there is the sense of agency that makes one aware that one is thinking and not doing something else (*Gallagher* 2012; *Proust* 2009): one has to take into account what is called agentive phenomenology (*Jansen* 2016). In such a phenomenology the performative or agentive aspects of mind processes are the main subject of the first-person inquiry: What is performed, what does the performer do, who is performing, what are the causes of this performance, what are the results of such processes and how do they differ from natural chains of events or thinking processes that just happen without agentive involvement of the subject having thoughts?

(5) Fifth, and last, one has to take into consideration seeing with the mind's activity, sometimes called "grasping" or the like (*Brown* 2004; *Chudnoff* 2015, pp. 39–40; *Pitt* 2004). This means that in thinking one is aware of universal conceptual relations that transcend individual consciousness but nevertheless appear within it (*Gutland* 2021; *Hopp* 2014). This is "not a matter of positing purely abstract ideality as metaphysically existent, but rather grasping or 'seeing' the universal in the individual" (*Froese* & *Gallagher* 2010, p. 89).

All these properties have been taken up in *Ziegler* & *Weger*, 2023; with this approach, ongoing research is well equipped to further the insight into aspects of mental actions within thinking processes.

## *7. Conclusion*

### *7.1. Experiments in thinking*

Thought experiments, in particular pure thought experiments, have their own experiential base. Their subject matter stems not from empirical observations but from conceptual evidence. They are concerned with abstract entities, namely with conceptual relations, structures, inferences, etc. We can feel secure in this conceptual realm and know how to navigate in it: we know the logical rules and can adhere to them or adjust them if new fields of inquiry necessitate such steps.

However, at first glance, we do not know much about how such thoughts are initiated, where they come from, how it feels to have such thoughts, who is the agent that initiates and pursues them, not to speak of the ontology of what these thoughts are about or what makes them appear in thinking consciousness. What we need to explore these matters further is the step from thought experiments, computer simulations, and mathematical thinking to *experiments in thinking*. We need to change the perspective from just doing these thinking activities to exploring the experiences while doing them. Hence, as has been suggested above, we need an experiential base to work with. The first step into this, before we can think of any grand theory or of formulating testable hypotheses, is *explorative experiments in thinking*.

### *7.2. Human agency in thinking*

If current research seeks to advance to the core issue, namely the initiation process of thinking actions, then there is no way around introspection and first-person accounts, which may be supplemented by other research perspectives but are indispensable. What better place to research questions of the autonomous self, of self and agency, of mental agency and mental action, etc., than within the scientific thinking activities that one feels confident in? In scientific thinking, and particularly in initiating experiments, we initiate the processes that happen. Do we not want to know how this initiation works? If we do, then the experiential material we gather by doing the types of pure thought experiments mentioned above and evaluating them using the methods of experiments in thinking, are highly relevant.

Research into this field of thinking actions may have far reaching consequences for understanding individual autonomy and the inner working of the conscious self (*Steiner* 1918; *Bandura* 2006). Thinking is at the center of all conscious actions. And if there is an individual autonomous

agency that initiates the thinking actions, then it is well worth exploring it beyond what is known by traditional psychological experiments.

### *7.3. Culture of experiments in thinking*

There is a well-established culture of experimenting in the natural, medical, and psychological sciences. It is an essential part of any scientific education in these fields to learn how to design and implement experiments. There is also a culture of thought experiments and of doing mathematics. However, we also need a culture and training of *experiments in thinking*. It is clear that just as naïve introspection (*Schwitzgebel* 2008) will not do, naïve experimentation in thinking will be of no help either. The author is convinced that such a culture and training (*Slagter et al.* 2007; 2011; *Ziegler* & *Weger*, 2019; 2023) might become part of the scientific repertoire that is needed for exploring thinking actions.

The important point here is that it does not suffice to just do, e.g., thought experiments or mathematics, but we must also notice and be aware of what happens while pursuing these actions, in order to note and avoid potentially confounding intrusions. Some of these issues are central to cognitive phenomenology (*Bayne* & *Montague* 2011; *Breyer* & *Gutland* 2016b) and the field of mental action and mental agency (*O'Brien* & *Soteriou* 2009). In spite of this, this type of thinking action is rarely taken into account; sometimes it is only referred to in passing (*Bayne* & *Montague* 2011, pp. 14–15), sometimes it is not mentioned at all (*Breyer* & *Gutland* 2016a), and sometimes it is explicitly excluded (*Chudnoff* 2015, p. 80). These issues have recently been taken up in a full-blown paper (*Ziegler* & *Weger* 2023).

With that, it is no surprise that the role of human agency in thinking and the issue of autonomy and self within thinking has not been explored much either (*Guillot* 2016; *Jansen* 2016; *Jorba* & *Moran* 2016). But this is exactly the subject this paper tries to bring up, in pondering over the initiation of scientific experiments and the performance of (pure) thought experiments. What is needed is a bridge from experiments in thinking to the question of an autonomous self, and even to the freedom of will in its strongest sense (*Chisholm* 2011; *Steiner* 1918), with all the controversy that this entails (*Klemm* 2010).

### *Acknowledgements*

Christopher Gutland and Ulrich Weger reviewed the manuscript and gave helpful comments on the argumentation and on the wording. Many thanks are going to Laura Liska concerning the improvement of my English and to Ruth Richter for her comments and suggestions towards a better understanding of my text.

*References*

*Allchin, D.* (2001): Error Types. Perspectives on Science 9(1), pp. 38–58.
*Anderson, F.* (2016): The Dynamic Phenomenology of Occurrent Thinking. Phenomenology and Mind 10, pp. 196–205.
*Anderson, F.* (2018): The Dynamic Phenomenology of Conscious, Occurrent Thinking: A First-Person Approach. United Kingdom.
*Antonini, S., Mariotti, M.A.* (2008): Indirect proof: What is specific to this way of proving? ZDM Mathematics Education 40(3), pp. 401–412.
*Bandura, A.* (2006): Toward a Psychology of Human Agency. Perspectives on Psychological Science 1(2), pp. 164–180.
*Bayne, T., Montague, M.* (2011): Cognitive Phenomenology: An Introduction. In: *Bayne, T., Montague, M.* (Eds.): Cognitive Phenomenology. Oxford, pp. 1–34.
*Brentano, F.* (1973): Psychologie vom empirischen Standpunkt (O. Kraus, Ed.). Hamburg.
*Breyer, T., Gutland, C.* (2016a): Introduction. In: *Breyer, T., Gutland, C.* (Eds.): Phenomenology of Thinking. Philosophical Investigations into the Character of Cognitive Experiences. New York, pp. 1–24.
*Breyer, T., Gutland, C.* (2016b): Phenomenology of Thinking: Philosophical Investigations into the Character of Cognitive Experiences. New York.
*Brown, J.R.* (2004): Peeking into Plato's Heaven. Philosophy of Science 71(5), pp. 1126–1138.
*Buchwald, J.Z.* (1998): Issues for the history of experimentation. In: *Heidelberger, M., Steinle, F.* (Eds.): Experimental Essays – Versuche zum Experiment. Baden-Baden, pp. 374–391.
*Buckareff, A.A.* (2005): How (Not) to Think About Mental Action. Philosophical Explorations 8(1), pp. 83–89.
*Buzzoni, M.* (2008): Thought Experiment in the Natural Sciences. Würzburg.
*Buzzoni, M.* (2011): On Mathematical Thought Experiments. Epistemologia 34, pp. 61–88.
*Carrier, M.* (1998): New experimentalism and the changing significance of experiments: On the shortcomings of an equipment-centered guide to history. In: *Heidelberger, M., Steinle, F.* (Eds.): Experimental Essays – Versuche zum Experiment. Baden-Baden, pp. 175–191.
*Chisholm, R.M.* (2011): Human freedom and the self. In*: Feinberg, J., Shafer-Landau, R.* (Eds.): Reason and Responsibility. Readings in Some Basic Problems of Philosophy. Boston MA, pp. 430–437.
*Chudnoff, E.* (2015): Cognitive Phenomenology. New York.
*Comte, A.* (1880): Einleitung in die positive Philosophie (Introduction to Positive Philosophy). Cours de philosophie positive, translated by Georg Heinrich Schneider. Leipzig.
*Dingler, H.* (1928): Das Experiment – Sein Wesen und seine Geschichte. München.

*Duhem, P.* (1998): Ziel und Struktur der physikalischen Theorien. Hamburg.
*Dutilh Novaes, C.* (2016): Reductio ad absurdum from a dialogical perspective. Philosophical Studies 173(10), pp. 2605–2628.
*Fehige, Y., Wiltsche, H.* (2013): The Body, Thought Experiments, and Phenomenology. In: *Frappier, M., Meynell, L., Brown, J.R.* (Eds.): Thought Experiments in Science, Philosophy, and the Arts. London, pp. 69–89.
*Fichte, J.G.* (1997): Grundlage der gesammten Wissenschaftslehre als Handschrift für seine Zuhörer. Introduction and register by W.G. Jacobs, Ed.; 4th ed. Hamburg.
*Froese, T., Gallagher, S.* (2010): Phenomenology and Artificial Life: Toward a Technological Supplementation of Phenomenological Methodology. Husserl Studies 26(2), pp. 83–106.
*Gallagher, S.* (2012): Multiple Aspects in the Sense of Agency. New Ideas in Psychology 30, pp. 15–31.
*Gallagher, S., Zahavi, D.* (2013): The Phenomenological Mind. 2nd ed. Hoboken.
*Glymour, C.* (2004): We believe in freedom of the will so that we can learn. Behavioral and Brain Sciences 27(05), pp. 661–662.
*Gooding, D.* (1990): Experiment and the making of meaning: Human agency in scientific observation and experiment. Dordrecht.
*Gooding, D.* (1992a): Putting agency back into experiment. In: *Pickering, A.* (Ed.): Science as Practice and Culture. Chicago-London, pp. 65–112.
*Gooding, D.* (1992b): The Procedural Turn; or, Why Do Thought Experiments Work? In: *Giere, R., Feigl, H.* (Eds.): Cognitive Models of Science. Minneapolis, pp. 45–76.
*Gooding, D.* (1992c): What is Experimental About Thought Experiments? PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association, pp. 280–290.
*Gooding, D., Pinch, T., Schaffer, S.* (1989): The Uses of Experiment: Studies in the Natural Sciences. Cambridge UK.
*Guillot, M.* (2016): Thinking of Oneself as the Thinker: The Concept of Self and the Phenomenology of Intellection. Philosophical Explorations 19(2), pp. 138–160.
*Gutland, C.* (2021): Psychological Consciousness of Non-psychological Contents: Aspects of a Phenomenology of Sensations and Thoughts. European Psychologist 26(2), pp. 73–84.
*Hacking, I.* (1983): Representing and Intervening: Introductory Topics in the Philosophy of Natural Science. Cambridge UK.
*Hacking, I.* (1992): The self-vindication of the laboratory sciences. In: *Pickering, A.* (Ed.): Science as Practice and Culture. Chicago, London, pp. 29–64.

*Heidelberger, M.* (1998): Die Erweiterung der Wirklichkeit im Experiment. In: *Heidelberger, M., Steinle, F.* (Eds.): Experimental Essays – Versuche zum Experiment. Baden-Baden, pp. 71–92.
*Heidelberger, M.* (2003): Theory-Ladenness and Scientific Instruments in Experimentation. In: *Radder, H.* (Ed.): The Philosophy of Scientific Experimentation. Pittsburgh, pp. 138–151.
*Heidelberger, M.* (2009): Das Experiment in den Wissenschaften. In: *Bartels, A., Stöckler, M.* (Eds.): Wissenschaftstheorie. Ein Studienbuch. 2nd ed. Paderborn, pp. 155–176.
*Heidelberger, M., Steinle, F.* (Eds.) (1998): Experimental Essays – Versuche zum Experiment. Baden-Baden.
*Hon, G.* (1998): If this be error: Probing experiment with error. In: *Heidelberger, M., Steinle, F.* (Eds.): Experimental Essays – Versuche zum Experiment. Baden-Baden, pp. 227–248.
*Hopp, W.* (2014): Experiments in Thought. Perspectives on Science 22(2), pp. 242–263.
*Janich, P.* (1998): Was macht experimentelle Resultate empiriehaltig? Die methodisch-kulturalistische Theorie des Experimentes. In: *Heidelberger, M., Steinle, F.* (Eds.): Experimental Essays – Versuche zum Experiment. Baden-Baden, pp. 93–112.
*Janich, P.* (2016): Experiment und Menschenbild. In: *Zeyer, K.* (Ed.): Hugo Dingler (1881–1954): Philosophie und Theorie des Experimentes. Regensburg, pp. 13–26.
*Jansen, J.* (2016): Kant's and Husserl's Agentive and Proprietary Accounts of Cognitive Phenomenology. Philosophical Explorations 19(2), pp. 161–172.
*Jorba, M., Moran, D.* (2016): Conscious Thinking and Cognitive Phenomenology: Topics, Views and Future Developments. Philosophical Explorations 19(2), pp. 95–113.
*Jorba, M., Vicente, A.* (2014): Cognitive Phenomenology, Access to Contents, and Inner Speech. Journal of Consciousness Studies 21 (9–10), pp. 74–99.
*Klemm, W.R.* (2010): Free will debates: Simple experiments are not so simple. Advances in Cognitive Psychology 6, pp. 47–65.
*Knobe, J., Buckwalter, W., Nichols, S., Robbins, P., Sarkissian, H., Sommers, T.* (2012): Experimental philosophy. Annual Review of Psychology 63, pp. 81–99.
*Knobe, J., Nichols, S.* (2017): Experimental Philosophy. In: *Zalta, E.N.* (Ed.): The Stanford Encyclopedia of Philosophy (Winter 2017). Metaphysics Research Lab. Stanford.
*Lange, R.* (2003): Technology as Basis and Object of Experimental Practices. In: *Radder, H.* (Ed.): The Philosophy of Scientific Experimentation. Pittsburgh, pp. 119–137.

*Lenhard, J.* (2018): Thought experiments and simulation experiments. In: *Stuart, M.T., Fehige, Y.J.H., Brown, J.R.* (Eds.): The Routledge Companion to Thought Experiments. London, pp. 484–497.
*Ludwig, K.* (2007): The Epistemology of Thought Experiments: First Person Versus Third Person Approaches. Midwest Studies in Philosophy 31(1), pp. 128–159.
*Ludwig, K.* (2018): Thought experiments and experimental philosophy. In: *Stuart, M.T., Fehige, Y.J.H., Brown, J.R.* (Eds.): The Routledge Companion to Thought Experiments. London, pp. 385–405.
*Mangan, B.* (2001): Sensation's Ghost: The Nonsensory Fringe of Consciousness. PSYCHE: An Interdisciplinary Journal of Research on Consciousness 7(18), pp. 1–44.
*Mele, A.* (1997): Agency and Mental Action. Philosophical Perspectives 11, pp. 231–249.
*Mele, A.* (2002): Goal-Directed Action: Teleological Explanations, Causal Theories, and Deviance. Noûs 34(14), pp. 279–300.
*Montague, M.* (2016): Cognitive Phenomenology and Conscious Thought. Phenomenology and the Cognitive Sciences 15(2), pp. 167–181.
*Nado, J.* (2016): Experimental Philosophy 2.0. Thought: A Journal of Philosophy 5(3), pp. 159–168.
*Nersessian, N.J.* (2018): Cognitive science, mental modeling, and thought experiments. In: *Stuart, M.T., Fehige, Y.J.H., Brown, J.R.* (Eds.): The Routledge Companion to Thought Experiments. London, pp. 309–326.
*O'Brien, L., Soteriou, M.* (Eds.). (2009): Mental Actions. Oxford.
*Petitmengin, C., Bitpol, M.* (2009): The Validity of First-Person Descriptions as Authenticity and Coherence. Journal of Consciousness Studies 16(10–12), pp. 363–404.
*Pitt, D.* (2004): The Phenomenology of Cognition or What Is It Like to Think That P? Philosophy and Phenomenological Research 69(1), pp. 1–36.
*Primas, H.* (1991): Necessary and sufficient conditions for an individual description of the measurement process. In: *Lahti, P., Mittelstaedt, P.* (Eds.): Symposium on the foundations of modern physics, Joensuu, Finland 1990. Singapore, pp. 332–346.
*Proust, J.* (2001): A Plea for Mental Acts. Synthese 129(1), pp. 105–128.
*Proust, J.* (2009): Is there a Sense of Agency for Thought? In: *O'Brien, L., Soteriou, M.* (Eds.): Mental Actions. Oxford, pp. 253–280.
*Putnam, H.* (1973): Meaning and Reference. Journal of Philosophy 70(19), pp. 699–711.
*Putnam, H.* (1975): The Meaning of "Meaning". Minnesota Studies in the Philosophy of Science 7, pp. 131–193.
*Radder, H.* (1998): Issus for a well-developed philosophy of scientific experimentation. In: *Heidelberger, M., Steinle, F.* (Eds.): Experimental Essays – Versuche zum Experiment. Baden-Baden, pp. 392–404.

*Radder, H.* (Ed.) (2003): The Philosophy of Scientific Experimentation. Pittsburgh.
*Radder, H.* (2009): The philosophy of scientific experimentation: A review. Automated Experimentation 1(2).
*Schickore, J.* (2005): "Through thousands of errors we reach the truth" – but how? On the epistemic roles of error in scientific practice. Studies in History and Philosophy of Science Part A 36(3), pp. 539–556.
*Schwitzgebel, E.* (2008): The Unreliability of Naive Introspection. Philosophical Review 117(2), pp. 245–273.
*Slagter, H.A., Davidson, R.J., Lutz, A.* (2011): Mental Training as a Tool in the Neuroscientific Study of Brain and Cognitive Plasticity. Frontiers in Human Neuroscience 5, 17.
*Slagter, H.A., Lutz, A., Greischar, L.L., Francis, A.D., Nieuwenhuis, S., Davis, J.M., Davidson, R.J.* (2007): Mental training affects distribution of limited brain resources. PLoS Biology 5(6), e138.
*Sorensen, R.A.* (1992): Thought experiments. New York.
*Sosa, E.* (2007): Experimental Philosophy and Philosophical Intuition. Philosophical Studies 132(1), pp. 99–107.
*Soteriou, M.* (2009): Mental Agency, Conscious Thinking, and Phenomenal Character. In: *O'Brien, L., Soteriou, M.* (Eds.): Mental Actions. Oxford, pp. 231–252.
*Spener, M.* (2011): Disagreement about cognitive phenomenology. In: *Bayne, T., Montague, M.* (Eds.): Cognitive Phenomenology. Oxford, pp. 268–284.
*Starikova, I., Giaquinto, M.* (2018): Thought experiments in mathematics. In: *Stuart, M.T., Fehige, Y.J.H., Brown, J.R.* (Eds.): The Routledge Companion to Thought Experiments. London, pp. 257–278.
*Steiner, R.* (1918): Die Philosophie der Freiheit (The Philosophy of Spiritual Activity). 2nd ed. Berlin; 17th ed., Dornach 2021, GA 4.
*Steinle, F.* (1997): Entering New Fields: Exploratory Uses of Experimentation. Philosophy of Science 64, pp. S65–S74.
*Steinle, F.* (1998): Exploratives vs. Theoriebestimmtes Experimentieren: Ampères erste Arbeiten zum Elektromagnetismus. In: *Heidelberger, M., Steinle, F.* (Eds.): Experimental Essays – Versuche zum Experiment. Baden-Baden, pp. 272–297.
*Steinle, F.* (2005): Explorative Experimente. Ampère, Faraday und die Ursprünge der Elektrodynamik. Stuttgart.
*Stich, S., Tobia, K.* (2018): Intuition and its critics. In: *Stuart, M.T., Fehige, Y.J.H., Brown, J.R.* (Eds.): The Routledge Companion to Thought Experiments. London, pp. 369–384.
*Stuart, M.T., Fehige, Y., Brown, J.R.* (2018): Thought Experiments: State of the Art. In: *Stuart, M.T., Fehige, Y.J.H., Brown, J.R.* (Eds.): The Routledge Companion to Thought Experiments. London, pp. 1–28.

*Tetens, H.* (1987): Experimentelle Erfahrung. Hamburg.
*Tye, M., Wright, B.* (2011): Is There a Phenomenology of Thought? In: *Bayne, T., Montague, M.* (Eds.): Cognitive Phenomenology. Oxford, pp. 285–325.
*Weger, U., Wagemann, J.* (2015a): The behavioral, experiential and conceptual dimensions of psychological phenomena: Body, soul and spirit. New Ideas in Psychology 39, pp. 23–33.
*Weger, U., Wagemann, J.* (2015b): The Challenges and Opportunities of First-Person Inquiry in Experimental Psychology. New Ideas in Psychology 36, pp. 38–49.
*Weger, U., Wagemann, J., Meyer, A.* (2018): Researching Mind Wandering from a First-Person Perspective. Applied Cognitive Psychology 32(3), pp. 298–306.
*Wiltsche, H.A.* (2018): Phenomenology and thought experiments. In: *Stuart, M.T., Fehige, Y.J.H., Brown, J.R.* (Eds.): The Routledge Companion to Thought Experiments. London, pp. 342–365.
*Wright, G.H. von* (1971): Explanation and Understanding. Ithaca.
*Ziegler R.* (2003): Teil II: Exploratives Experimentieren, ideales Experiment und konditionaler Determinismus. Elemente d. N. 79, pp. 22–50.
*Ziegler, R., Weger, U.* (2018): First-Person Experiments in Thinking. European Psychologist 23(3), pp. 189–205.
*Ziegler, R., Weger, U.* (2019): Exploring Conceptual Thinking and Pure Concepts from a First Person Perspective. Phenomenology and the Cognitive Sciences 18(5), pp. 947–972.
*Ziegler, R., Weger, U.* (2023): Thinking Action as a Performative and Participative Mental Awareness. Frontiers in Psychology – Consciousness Research 14:901678. DOI: 10.3389/fpsyg.2023.901678

*Renatus Ziegler*
*Hügelweg 29*
*CH – 4143 Dornach*
*r.ziegler@reinesdenken.ch*