“Platonic” thought experiments: how on earth?

Rafal Urbaniak
Department of Philosophy, Gdansk University, Poland
Centre for Logic and Philosophy, Ghent University, Belgium
rfl.urbaniak@gmail.com

May 13, 2011

Abstract. Brown (1991a,b, 2004, 2008) argues that thought experiments (TE) in science cannot be arguments and cannot even be represented by arguments. He rest his case on examples of TEs which proceed through a contradiction to reach a positive resolution (Brown calls such TEs “platonic”). This, supposedly, makes it impossible to represent them as arguments for logical reasons: there is no logic that can adequately model such phenomena. (Brown further argues that this being the case, “platonic” TEs provide us with irreducible insight into the abstract realm of laws of nature). I argue against this approach by describing how “platonic” TEs can be modeled within the logical framework of adaptive proofs for prioritized consequence operations. To show how this mundane apparatus works, I use it to reconstruct one of the key examples used by Brown, Galileo’s TE involving falling bodies.

Keywords: thought experiments, platonism, prioritized consequence, entrenchment, Galileo, adaptive logics

1. Philosophical Motivations

Empiricism about thought experiments (TEs), especially as they occur in the history of science, is the view that TEs are nothing but “ordinary argumentation that is disguised in a vivid pictorial or narrative form” (Norton, 2004b, 45). This view, represented for instance by (Norton, 1996; Norton, 2004a; Norton, 2004b), is rejected by (Brown, 1991a; Brown, 1991b; Brown, 2004; Brown, 2008), who embraces the opposite view dubbed Platonism about TEs and insists that TEs provide us with insight into the abstract realm of laws of nature, resulting in what he calls “*a priori* (though still fallible) knowledge of nature” (Brown, 2008).

Brown’s argumentative strategy can be summarized as follows:

– He argues that there are TEs which cannot be reconstructed as arguments and nevertheless provide us with new information.

– It is rather uncontroversial that such TEs are not empirical observations.

- He suggests that the only way they can provide new information while being neither arguments nor empirical observations is if they are acts of perceiving the abstract realm of laws of nature (and because he postulates the existence of such a realm, his view is called ‘platonism’).

While the last move is debatable, I will not discuss it in this paper. Instead, I will focus on the first step and argue that the argument can be blocked already at that step.

Brown rests his case for his first claim on the existence of what he calls “platonic” TEs. Those are TEs which not only perform a destructive task by motivating the rejection of one of the initially accepted beliefs, but also lead to the correct resolution of the encountered anomaly. Brown claims that if such arguments could be constructed as arguments, there would be a logic which would capture the inferences involved. By contraposition, he suggests that since there is no logic which can handle the dynamics of platonic TEs, they cannot be interpreted as arguments.

While the move from the claim that TEs cannot be represented by arguments to platonism about the laws of physics might seem hasty, Brown has a point in saying that so far no formalized reconstruction of platonic TEs does justice to their dynamics. How exactly can we start a TE, reach a contradiction and yet, within one and the same TE, end up with a more or less acceptable conclusion, all the time using a single sensible logical framework?

The goal of this paper is to describe a formal logical framework within which the thought-experimental moves that Brown finds essential for his argument can be modeled. Thus, the intention is to undermine his approach to TEs by answering the logical challenge he posed.

2. Galileo on falling bodies

A classical example of what Brown calls “platonic” TEs is Galileo’s celebrated TE concerned with the speed of falling bodies of different weights. Suppose you’re working within the Aristotelian framework. Then, the difference in speed between two freely falling bodies is proportional to the difference in their weight. Imagine a heavy cannon ball l and a really tiny musket ball made of the same material s falling from a certain fixed height.

According to the Aristotelian assumption, l will be falling faster than s. Imagine you join l and s together: since s is slower than l, it will slow l down, so the speed of $l+s$ (let’s denote the operation of joining bodies by the addition symbol) will be lower than the speed of l. On the other hand, $l + s$ is heavier than l itself, and that being the case, $l + s$ has to fall faster than l on its own. So, we have a contradiction (given the assumption that one thing cannot fall simultaneously slower and faster than another). This is the destructive part of the TE. Further considerations (variously reconstructed by different interpreters like (Schrenk, 2004), (Brown, 1991a), or (Gendler, 1998)) lead to the conclusion that in fact, those bodies have to be falling at the same rate.

(Brown, 1991a, 77-78) argues:

> There have been no new empirical data [...] the transition from the old to the new theory [...] is not readily explained in terms of empirical input unless there is new empirical input. Galileo's new theory is not logically deduced from old data. Nor is it any kind of logical truth. A second way of making new discoveries [...] is by deducing them from old data. Norton holds such a view when he claims that a thought experiment is really an argument. [...Will this view] account for those [thought experiments] I call platonic? I think not. *The premisses of such an argument could include all the data that went into Aristotle's theory.* [emphasis mine, RU] From this Galileo derived a contradiction (So far, so good; we have a straightforward argument which satisfies Norton's account.) But can we derive Galileo's theory that all bodies fall at the same rate from these same premises? Well, in one sense, yes, since we can derive anything from a contradiction; but this hardly seems fair. What's more, whatever we can derive from these premisses is immediately questionable since, on the basis of the contradiction, we now consider our belief in the premisses rightly to be undermined.

3. Straightforward reconstructions of Galileo's TE

Norton doesn't explicitly address Brown's *logical* concern (how can the dynamic aspect of platonic TEs be modeled by logical means?). In general, he doesn't seem to pay too much attention to logical details. He thinks that familiar logics (he doesn't really specify which those are) will suffice. To support this view, he gives what he calls an "evolutionary argument":

> I think there are some reasons to believe that no new, exotic logic is called for. In outlining the general notion of logic above, I recalled the evolutionary character of the logic literature in recent times. New inferential practices create new niches and new logics evolve to fill them. Now the activity of thought experimenting in science was identified and discussed prominently a century ago by Mach

> (1906) and thought experiments have been used in science actively for many centuries more. So logicians and philosophers interested in science have had ample opportunity to identify any new logic that may be introduced by thought experimentation in science. So my presumption is that any such logic has already been identified, in so far as it would be of use in the generation and justification of scientific results. I do not expect thought experiments to require logics not already in the standard repertoire. This is, of course, not a decisive argument. Perhaps the logicians have just been lazy or blind. It does suggest, however, that it will prove difficult to extract a new logic from thought experiments of relevance to their scientific outcomes – else it would already have been done! (Norton, 2004b, 54-55)

To gain some perspective on this argument, consider the following "evolutionary" argument. Norton, among other things, works on philosophy of relativity. Now, relativity theory has been around, pretty much, since the time when Mach wrote about TEs. So philosophers interested in science have had ample opportunity to identify and solve any philosophical issue that may be introduced by relativity theory. So, in this field, any philosophically interesting claim has already been made and any philosophically interesting argument has been given, and Norton's work in philosophy or relativity is redundant. Unless, of course, philosophers of science since the discovery of relativity theory have just been lazy or blind.

In fact, Norton's reconstruction of Galileo's TE (Norton, 1996, 341-3) is logically quite straightforward. To obtain the destructive part of the argument, he identifies all the assumptions needed to derive a contradiction from the Aristotelian assumption, takes them as assumptions of the argument, uses the Aristotelian assumption for *reductio*, and derives a contradiction. Then, he adds one more assumption (that the speed of falling bodies depends only on their weights) and argues that the claim that the objects fall with the same speed follows.[1] Next, he

[1] Initially, it might be unclear why this is supposed to constitute an addition: the Aristotelian assumption entails it. To see why, let's take a look at a very simplified representation. Say '$W(x)$' and '$S(x)$' stand for the weight and speed of x respectively. Then, the assumption that the speed depends only on weight is

$$\text{[DepOn]} \quad \neg\exists_{x,y}[W(x) = W(y) \wedge S(x) \neq S(y)]$$

and the Aristotelian assumption reads

$$\text{[Ar]} \quad \forall x, y\, [W(x) > W(y) \equiv S(x) > S(y)]$$

Suppose [DepOn] fails while [Ar] holds. So for some a, b: $W(a) = W(b)$ but $S(a) \neq S(b)$. Then, either $S(a) > S(b)$ or $S(b) > S(a)$. In the first case, by [Ar], $W(a) > W(b)$. In the second case, by [Ar], $W(b) > W(a)$. Either way we contradict the explains that this additional assumption actually was not acceptable in the original context of the thought experiment, and concludes that "this final step now looks more like a clumsy fudge or a stumble than a leap into the Platonic world of laws." (Norton, 1996, 345)

Another construal of the Galilean argument has been given by Graham Priest.[2] The argument starts with the assumption that either A will be falling down faster than b, or b will be falling down faster than A, or they will be falling with the same speed. Then, two *reductio* arguments are employed to exclude the first two options, thus leaving us with the only remaining solution.

4. Weak points of straightforward reconstructions

Whether we are to take the straightforward reconstructions of the sort mentioned above to be successful clearly depends on what we want them to accomplish.

Sure, important chunks of processes in those TEs behave like those straightforward arguments. But given our current motivations, we need to measure their success against the challenge posed by Brown. For him, the Galilean TE starts with the initial acceptance of the Aristotelian assumption (see quote on p. 3), proceeds through an actual contradiction and reaches the resolution.

The Aristotelian assumption is neither accepted unconditionally (if it were, it could not be overthrown by further considerations), nor is it assumed without genuine acceptance, merely for *reductio* (for Galileo's Aristotelian opponent really accepts it). It rather seems that the assumption is accepted *defeasibly*, so that its acceptance is open to revision (and in fact, being revised during the process).

These aspects are not modeled by Norton's set-up in which the Aristotelian claim is merely assumed for *reductio*, the obtained contradiction does not collide with the agent's initial beliefs and further steps towards the final conclusion are just taken to be clumsy.[3]

assumption that $W(a) = W(b)$. But one has to remember that the assumption is added to the set of initial premises *minus* the Aristotelian assumption. And indeed, [DepOn] is weaker than [Ar].

[2] In a verbal discussion following my talk at the *Logic, Reasoning Rationality 2010* conference organized by Ghent University in Belgium.

[3] Observe that in Norton's reconstruction some heavy-lifting is done by the choice of what is taken as a mere assumption and what is taken as a *reductio* assumption of the proof. (Schrenk, 2004) reconstructs the destructive argument in more detail than Norton, and suggests that logically speaking, it does not unambiguously lead to the rejection of the Aristotelian assumption (the rejection of any of the premises

Norton's description, instead of a single formally reconstructed argument, involves the interplay of a few arguments best described in meta-language rather than modeled in a formal system (see below). Moreover, Norton merely uses the Aristotelian assumption for *reductio* without representing its initial defeasible acceptance by the Aristotelian. By giving an account which does not employ a single formally reconstructed argument and does not model defeasible acceptance, Norton already makes an unnecessary concession to Brown.

Another option is to stay classical, but instead of formally reconstructing the TE, to tell a story in meta-language from an external perspective. This would involve saying that at different times people involved used slightly different assumptions which they revised for good reasons. While this approach makes it look more like a single argument, it does not really explain the logical mechanism underlying the revisions. In this sense, this strategy fails to answer Brown's challenge, who asked about the underlying *logic*. To satisfy this requirement, I take it, a formalized reconstruction satisfying all of the above described desiderata has to be provided.

If we want to be able to formally reconstruct a TE as a single argument, something else than classical logic is needed. Of course, one is free to insist that apart from the classical reconstructions there is no interesting story to be told and to deny the need to model other aspects of the rational processes in question formally. This however means that one is not playing the same game as Brown anymore. He demands a unifying formal account which captures also some aspects which straightforward reconstructions fail to capture. While claiming that this challenge doesn't have to be met is one way to respond, it is unlikely to push the debate forward. The question whether there are sensible logical systems which capture what Brown requires them to capture still remains.

Brown's qualms aside, reconstructing TEs in physics as arguments has some independent virtues. TEs can err (for a few nice examples, see (Norton, 2004b)) and the error can stem from what we tacitly accept in the TE. Formalization allows us to see all the assumptions involved, and this makes it easier to assess them.

5. What to do?

(Gendler, 1998) emphasizes that logically speaking, there are at least four non-trivial ways out for the Aristotelian, when faced with the

employed in the argument would suffice, if one were guided only by the desire to avoid contradiction).

destructive part of Galileo's TE, and that for the right outcome a certain prior preference on the premises involved is needed.[4] Gendler's reconstruction involves such a preference and belief revision. It is quite natural, but it's described informally and the request for a *logic* which underlies the reasoning involved still has to be answered.

Thus, it would be useful to have a formal logic *with proof theory* which adequately represents reasoning of the "platonic" sort. What *desiderata* should it satisfy? The system should be able to keep track of preferences between premises involved, because the lack of prioritization seems to be the key shortcoming of the logic employed in straightforward reconstructions. It also should be able to model the rejection of the least entrenched assumptions upon encountering a contradiction, without running into logical triviality. Finally, after encountering a contradiction and rejecting one of the premises, without the logical explosion, the system should allow one to use the remaining premises in a sensible manner to derive the resolution.

I will argue that prioritized adaptive logics with their proof theory satisfy these requirements. First, I will explain what adaptive logics and prioritized adaptive logics are. Then, I will use a prioritized adaptive logic to reconstruct the destructive part of Galileo's TE. Next, I will show that there is an assumption that allows one to derive, without any "clumsy fudges and stumbling," that the two objects involved in the TE will be falling at the same rate.[5] Both the destructive and the constructive processes will be modeled within one and the same argument, governed by one formal logical system.

On the approach proposed in this paper, the adaptive formal framework will be argued to be a convenient tool for capturing how TEs (to borrow a phrase from Kuhn) "assist in the elimination of prior confusion by forcing the scientist to recognize contradictions that had been inherent in his way of thinking from the start" even though "the elimination of existing confusion does not seem to demand additional empirical data" (Kuhn, 1977, 242). The philosophical upshot will be that Brown's argument that some sort of platonic insight must be involved because logic can't handle further arguments once a contradiction is derived, fails.

[4] She also has a fascinating epistemological story to tell about how the preference is discovered (Gendler, 2004; Gendler, 2007), but those issues are too far from my current concern.

[5] It is enough to assume that if the two objects involved are made of the same material and are approximately of the same shape then the lighter one will not fall faster than the heavier one (this assumption is weaker than any of the assumptions suggested by Norton or Gendler).

6. A simple adaptive logic

Adaptive logics with their dynamic proofs are quite complicated animals. Before I move to dynamic proofs for prioritized consequence relations, let us take a look at an adaptive system devised to handle some simple arguments about expectancies.

Adaptive logics are so called because they adapt themselves to the premises they are applied to: the applicability of some rules or steps depends on the premises and on what conclusions have been derived at a given stage of a proof.[6] Roughly speaking, while reasoning adaptively we use rules of two simpler logics (called the 'lower limit logic,' **LLL** and the 'upper limit logic,' **ULL**, **ULL** being a strengthening of **LLL**). The specific rules of the stronger logic are applied in proofs conditionally upon the normal behavior of certain formulas (that is, upon the falsehood of formulas whose truth is to be avoided if possible – they're often called *abnormalities*), and if further in the proof it turns out that those formulas do not behave normally, steps depending on their normal behavior are retracted. Given a **ULL** and **LLL**, different choices of sets of abnormalities lead to different adaptive logics.

This is all very general and hand-wavy, but examples which I will soon give should make it clear how various adaptive logics fit this general profile. A mathematically precise and general definition of the so-called *standard format* of adaptive logics is available (Batens, 2007).[7] However, since it involves various technicalities not needed for current considerations, I will avoid this level of detail and rather use examples to allow the reader to understand enough of the formal systems to grasp their applicability to the philosophical issue at hand. The standard format also provides adaptive logics with a unified model theory: once an adaptive logic falls under the standard format, it has an array of meta-theoretic properties (like soundness and completeness). Since, however, I am not interested in model theory in this paper, such issues will be ignored. What will matter is the description of the consequence operation, the corresponding proof theory and the applicability of the logic to the philosophical problems we are interested in.

The fact that in an adaptive proof some steps can become retracted once new information is derived allows for the representation of arguments which may be doubly (externally and internally) dynamic.

[6] Some basic papers about adaptive logics are (Batens, 1995; Batens, 2004; Batens, 2007). For more references, see the website of the Centre for Logic and Philosophy of Science at Ghent University, http://logica.ugent.be/centrum/writings/.

[7] There are adaptive logics that do not fit the standard format, but the working (and confirmed by numerous cases) hypothesis of Ghent research group is that all adaptive logics are equivalent to adaptive logics in the standard format.

Externally dynamic, because most of adaptive logics are *nonmonotonic*: once our premise set is extended by new input, we might have to retract some of our previous conclusions if the new information makes some steps unreliable. Adaptive logics are also internally dynamic because even with the same premise set, it may turn out that a conclusion which is **ULL**-derived is no longer reliable once at some later stage of the proof a problematic formula becomes **LLL**-derived. Now, we'll take a look at a simple example of an adaptive logic (I will restrict myself to the propositional case).

In this simple adaptive logic **LLL** is the modal logic **T** for the standard modal propositional language built from a countable supply of propositional variables ($p, q, r, p_1, q_2, r_2, \ldots$), negation ($\neg$), conjunction ($\wedge$), disjunction ($\vee$), implication ($\rightarrow$) and modal operators ($\Box$ and $\Diamond$). **T** is axiomatized by classical propositional logic (**CL**) strengthened with all substitutions of axioms **K** and **T**

$$\begin{array}{ll} \mathbf{K} & \Box(\phi \rightarrow \psi) \rightarrow (\Box\phi \rightarrow \Box\psi) \\ \mathbf{T} & \Box\phi \rightarrow \phi \end{array}$$

and the necessitation rule which from the fact that ϕ is a theorem ($\vdash \phi$) allows to infer that its necessitation is a theorem ($\vdash \Box\phi$).

In what follows we will need the fact that all substitutions of the following are theorems of **T**.[8]

$$\begin{array}{ll} \mathbf{T1} & \Diamond\phi \rightarrow (\phi \vee (\Diamond\phi \wedge \neg\phi)) \\ \mathbf{T2} & \Diamond(\phi \wedge \psi) \rightarrow \Diamond\phi \end{array}$$

A slightly unusual feature of our interpretation of the modal language is that we read $\Diamond\phi$ (where ϕ is a non-modal formula) as 'it is expected that ϕ'. This reading indicates where the dynamic aspect comes in: we want to accept expectancies insofar as they do not contradict the data, and to rectract conclusions which relied on expectancies which later on turned out to contradict the data. This means, we want to defeasibly assume that as many expectancies are true as we consistently can assume to be true: we want to reject as many formulas of the form $\Diamond\phi\wedge\neg\phi$ (where ϕ is non-modal) as we can. I will call such formulas *abnormalities* and abbreviate them sometimes as $!\phi$. Once we define abnormalities this way, the upper limit logic is just **T** strengthened with the assumption that all abnormalities are false.

A dynamic proof is a sequence of lines which consist of four components: a *line number*, a *formula* (which we will call the formula *of*

[8] **T1** is a theorem of propositional logic. **T1** is a trivial theorem of logic **K** and all its extensions.

that line), a *justification* for that formula, and a possibly empty set of formulas.[9] Besides, each line can be marked (marks can come and go as the proof progresses). If a line is at some point marked, it means that the formula of this line is not considered derived at that stage. The first three components are rather clear. Conditions and marking require some more attention.

There are three rules for proofs from a premise set Γ. The first rule, Prem allows one to introduce any premise $\phi \in \Gamma$ with the empty set in the *conditions* column. That is if $\phi \in \Gamma$, it allows to infer:

$$(n) \quad \phi \quad \textsc{Prem} \quad \emptyset$$

where n is an appropriate line number.

The second rule, Ru says that if we have proven ϕ_1 on Δ_1 (that is, we have Δ_1 among the conditions of a line where ϕ_1 is the formula), ϕ_2 on $\Delta_2, \ldots,$ and ϕ_n on Δ_n, and if ψ can be **LLL**-derived (in the present case, **T**-derived) from $\phi_1, \ldots, \phi_n$, we can introduce ψ as relying on the normal behavior of $\Delta_1 \cup \Delta_2 \cup \cdots \cup \Delta_n$. That is, if

$$\{\phi_1, \ldots, \phi_n\} \vdash_{\mathbf{T}} \psi$$

then

from			
(i)	ϕ_1		Δ_1
(j)	ϕ_2		Δ_2
		$\vdots$	
(k)	ϕ_n		Δ_n
derive			
(l)	ϕ	Ru: i, j, …, k	$\Delta_1 \cup \Delta_2 \cdots \cup \Delta_n$

where $i < j < k < l$ are appropriate line numbers. That is, we can add **T**-consequences relying on nothing more and nothing less than the union of those sets on which the premises depended.

The third rule, Rc, is based on the following idea. If from Γ we can **T**-derive that either ψ is true or one of the formulas in a set of abnormalities Δ is true, we can conclude that ψ follows in our adaptive logic from Γ on the defeasible assumption that formulas in Δ are false. If Δ is a finite set of abnormalities, let us call the disjunction of its

[9] Often, these are abnormalities upon the assumption of whose falsehood the formula is derived. Not necessarily so in the so-called direct proof theories. The column containing such sets will be called a dependence column or a conditions column.

members '$Dab(\Delta)$'. If for some finite set of abnormalities Θ:

$$\{\phi_1, \ldots, \phi_n\} \vdash_{\mathbf{T}} \psi \vee Dab(\Theta)$$

then

from			
(i)	ϕ_1		Δ_1
(j)	ϕ_2		Δ_2
		$\vdots$	
(k)	ϕ_n		Δ_n
Derive			
(l)	ψ	Rc: i, j, ..., k	$\Delta_1 \cup \Delta_2 \cdots \cup \Delta_n \cup \Theta$

That is, if either ψ is true or one of the abnormalities in Θ is true, then ψ holds, as long as formulas in Θ behave normally (=are false). If $\psi \vee Dab(\Theta)$ is derived from $\phi_1, \ldots \phi_n$ by means of Prem and Ru only, then Rc allows to move $Dab(\Theta)$ into the conditions column and add it to the union of dependencies of $\phi_1, \ldots, \phi_n$.

In particular, thanks to **T1**, if ϕ is non-modal, from $\Diamond\phi$ we can **T**-derive $\phi \vee !\phi$, and then Rc allows us to derive ϕ from $\Diamond\phi$ on the condition $!\phi$. Another way to think about this is to consider the (defeasible) assumption that a certain abnormality is false: $\neg !\phi$. This means $\neg(\Diamond\phi \wedge \neg\phi)$ and is equivalent to $\Diamond\phi \rightarrow \phi$. Thus, our adaptive logic allows us (defeasibly) to drop single diamonds in front of non-modal formulas.

Our task in a proof is not only to derive formulas from premises but also to recognize those steps which cannot be trusted. What do we mean by this? At the first (and not completely correct) stab, a step is unreliable if it depends on the falsehood of an abnormality which as it turns out **T**-follows from the premises.

This is quite close. There is a complication, though. A premise set may **T**-prove $Dab(\Delta)$ without proving any element of Δ separately. In such a case we learn that at least one of the formulas involved in the dependence column doesn't behave normally, but we have no idea which disjunct is responsible. To deal with this issue, we need something more.

A proof, as we conduct it, proceeds in stages. Every application of a rule carries us to the next stage. A formula $Dab(\Delta)$ is a *minimal Dab-formula* of a proof at a stage s iff $Dab(\Delta)$ occurs in the proof at a line with the condition $\emptyset$ (that is, if it's derived from the premises by means of **T** only) and for no $\Delta' \subset \Delta$ (that is, for no Δ' which is a proper subset of Δ) the proof at stage s contains a line with $Dab(\Delta')$ derived on condition $\emptyset$.

We need to define which lines of an adaptive proof are marked at which stages (intuitively, a marking symbol next to a line means the formula in that line is not derived at a given stage of the proof). One of the simplest plausible marking definitions in this context[10] is the one based on *reliability*. On this definition, at a given stage a line is marked (as unreliable) if it depends on a set of abnormalities Δ, and at that stage some member of Δ is a disjunct in a minimal disjunction of abnormalities **T**-derived from the premises.

The intuitive reason why we are interested in *minimal Dab*-formulas of a proof instead of just *any* proven *Dab*-formulas whatsoever is this. We want *Dab*-formulas to help us discover those abnormalities whose normal behavior is not to be expected. The fact that $Dab(\Delta)$ has been **T**-proven from the premises tells us only that at least one member of Δ has to be true if the premises are to be true. However, we want to assume that as many abnormalities are false as possible and we take any abnormality to be false unless compelled to do otherwise. So, if we know that both $Dab(\Delta)$ and $Dab(\Delta')$ are **T**-derived from our premises, but also that $\Delta' \subset \Delta$, we know that we do not have to blame any member of $\Delta \setminus \Delta'$. If we want to accept as few abnormalities as possible, it will suffice to assume that it is the members of Δ' that are not reliable.

Hence, we first define $U_s(\Gamma)$ to be the union of all Δs that are constituents of those minimal *Dab*-formulas that have been derived so far from Γ at stage s. Then, a line in a proof is marked at stage s if it depends on the normal behavior of Θ, and yet at least one member of Θ is a member of $U_s(\Gamma)$. That is, if a certain line contains a certain set in the dependence column, and yet at least one member of this set is among those abnormalities on whose normal behavior we can't rely, the formula of that line is not considered derived. As the proof proceeds, the list of minimal *Dab*-formulas and $U_s(\Gamma)$ might change. If they do, certain line marks might come or go.

Since sometimes an unmarked line becomes marked later in the proof, the fact that a line is derived and unmarked at a certain stage does not mean that it really follows from the premises. (Similarly, the fact that a line is marked at a certain stage, doesn't mean that it won't become unmarked at some later stage and that it doesn't follow from the premises). Hence we also need the notion of *final derivability*. A formula is finally derived in a proof if it is derived in an unmarked line of that proof at a finite stage and also any extension of the proof in which it becomes marked can always be itself extended into a proof where it is unmarked.

[10] A variety of marking definitions, depending on the goal of a logic is available (Batens, 2007). You will also see another marking definition in section 8.

Suppose our premise set is $\Gamma = \{\Diamond(p \wedge q), \neg p \vee \neg q, \neg q\}$. Consider the following adaptive proof (I present it in a slightly condensed form, superscripting marks with line numbers to indicate when those marks appear and disappear).

(1)	$\Diamond(p \wedge q)$	Prem	$\emptyset$	
(2)	$\Diamond p$	Ru: 1, **T2**	$\emptyset$	
(3)	$\Diamond q$	Ru: 1, **T2**	$\emptyset$	
(4)	$p \vee (\Diamond p \wedge \neg p)$	Ru: 2, **T1**	$\emptyset$	
(5)	$q \vee (\Diamond q \wedge \neg q)$	Ru: 3, **T1**	$\emptyset$	
(6)	p	Rc: 4	$\{!p\}$	$\curlyvee^{9,10}$
(7)	q	Rc: 5	$\{!q\}$	$\curlyvee^{9,10,11}$
(8)	$\neg p \vee \neg q$	Prem	$\emptyset$	
(9)	$(\Diamond p \wedge \neg p) \vee (\Diamond q \wedge \neg q)$	Ru: 2, 3, 8	$\emptyset$	
(10)	$\neg q$	Prem	$\emptyset$	
(11)	$\Diamond q \wedge \neg q$	Ru: 3, 10	$\emptyset$	

In line (1) we just introduce a premise. In lines (2) and (3) we apply **T2** twice to distribute $\Diamond$ over a conjunction. Lines (4) and (5) follow by **T1**, and they are interesting because their second disjuncts are abnormalities. Given the fact that lines (4) and (5) **T**-follow from the premises, we are allowed to conditionalize on the normal behavior of involved abnormalities, thus introducing p conditionally on the falsity of $\Diamond p \wedge \neg p$ and q conditionally on the falsity of $\Diamond q \wedge \neg q$ in lines (6) and (7). In line (8) we introduce another premise, which (together with lines (2) and (3)) **CL**-entails line (9). At this point, our $U_9(\Gamma) = \{!p, !q\}$ (its elements occur in a – so far – minimal derived disjuction of abnormalities). This means the lines marked right after the introduction of line (9) are (6) and (7), because each of them depends on an abnormality which is in U_9. In line (10) we introduce the last premise[11] which classically entails line (11). But once line (11) is derived, the formula from line (9) no longer is a minimal Dab-formula, and indeed U_{11} for our proof simply is $\{!q\}$. This means line (6) is at this stage no longer suspect and is unmarked, while line (7) still relies on a formula in U_{11} and remains marked. It is also clear that (6) contains a finally derived formula.

This should give the reader a sufficient grasp of what dynamic proofs look like. Now, we can introduce another important element of

[11] It is not required that all premises are introduced in the beginning of the proof. On the other hand, the reader certainly can see that I introduce premises in a somewhat artificial order, but I do it to be able to indicate a few different phenomena in a single simple proof.

our formal reconstruction of Galileo's TE: the prioritized consequence operation **ND** and its adaptive proof theory.

7. Non-defeated prioritized consequence operation

Say we are dealing with *prioritized belief bases*. Such a base Σ is identified with a finite tuple of consistent belief levels Σ_i, which contain well-formed sentences of a given language: $\langle\Sigma_1, \Sigma_2, \ldots, \Sigma_n\rangle$. The basic intuition here is that the lower the subscript is, the more important the assumptions that belong to this set are. Note the assumption that each level is consistent. This doesn't mean that one's beliefs in general have to be consistent. The assumption here is weaker: it consists in the restriction that if a belief set is inconsistent, it is not a set of beliefs of the same level of entrenchment. To some extent, this is an idealization, for it excludes cases where we have equal support for opposing conclusions. On the other hand, if this does happen, we rationally should do our best to reassess the reasons we have and to either undermine one of them or to strengthen the other. Since dealing with these issues would lead us far beyond our considerations, this remark should suffice for now.[12]

As it turns out, there are many different ways we can delineate systematic inferential practices that one might use to deal with prioritized beliefs when some of them lead to a contradiction.[13] I will employ the non-defeated consequence operation (**ND**), not only because I find it intuitive, but also because it is rather conservative compared to other approaches (Benferhat et al., 1997), so whatever follows by **ND**, follows also from the perspective of most of other approaches to prioritized reasoning.[14]

First, we say that Δ is a *maximal consistent subset* of Γ iff $\Delta \subseteq \Gamma$, Δ is consistent and for any $\phi \in \Gamma \setminus \Delta$ the result of adding ϕ to Δ is inconsistent. Then we say that a formula ϕ is free in Γ ($\phi \in F(\Gamma)$) iff ϕ belongs to every maximal consistent subset of Γ. To see how the selection of free formulas works, consider a few examples.

Example 1. Suppose $\Gamma = \{p, \neg p, q\}$. Intuitively, p and $\neg p$ are problematic and q is innocent. Γ has two maximal consistent subsets: $MC_1 =$

[12] Sometimes, I will be sloppy and talk about a formula belonging to Σ instead of it belonging to some Σ_i which belongs to Σ, but I think this will save space and won't cause any important ambiguity.

[13] For a survey, see (Benferhat et al., 1998) and (Verhoeven, 2003).

[14] Choosing logic for a goal we often are more concerned with satisficing than optimizing (Herbert, 2008).

$\{p, q\}$ and $MC_2 = \{\neg p, q\}$. Neither p nor $\neg p$ belongs to both of them, so $p, \neg p \notin F(\Gamma)$. However, $q \in MC_1$ and $q \in MC_2$, so $q \in F(\Gamma)$. $\dashv$

Example 2. Say $\Gamma = \{\neg p, p \vee q, \neg q\}$. There are three maximal consistent subsets of Γ: $\{\neg p, p \vee q\}$, $\{\neg q, p \vee q\}$ and $\{\neg p, \neg q\}$. None of the formulas from Γ belongs to all of them, so $F(\Gamma) = \emptyset$. Indeed, intuitively speaking, each of those formulas can play an essential role in deriving a contradiction, so each of them is suspicious.$\dashv$

Example 3. Extend the set from the previous example to $\Gamma = \{\neg p, p \vee q, \neg q, r\}$. Intuitively, r has nothing to do with the fact that Γ is inconsistent. And indeed, there are three maximal consistent subsets of Γ: $\{\neg p, p \vee q, r\}$, $\{\neg q, p \vee q, r\}$ and $\{\neg p, \neg q, r\}$, each of them contains r, but none of the other formulas belongs to each of them. So $r \in F(\Gamma)$, but $\neg p, p \vee q, \neg q \notin F(\Gamma)$. $\dashv$

The *dominant subset* of Σ is $\Sigma^\star = F(\Sigma_1) \cup F(\Sigma_1 \cup \Sigma_2) \cup \cdots \cup F(\Sigma_1 \cup \cdots \cup \Sigma_n)$. This is supposed to be the set of those formulas which are not suspicious, built with the preference to more entrenched premises.

Example 4. To observe how this encodes the preference relation, consider a very simple case where $\Sigma = \langle \Sigma_1, \Sigma_2 \rangle$, $\Sigma_1 = \{p\}$ and $\Sigma_2 = \{\neg p\}$. There is exactly one maximally consistent subset of Σ_1, namely $\{p\}$ itself. Thus, $p \in F(\Sigma_1)$ because $F(\Sigma_1) = \{p\}$. $\Sigma_1 \cup \Sigma_2$ has two maximally consistent subsets: $\{p\}$ and $\{\neg p\}$. Since no formula is in both of them, $F(\Sigma_1 \cup \Sigma_2) = \emptyset$. But this means that $\Sigma^\star = F(\Sigma_1) \cup F(\Sigma_1 \cup \Sigma_2) = F(\Sigma_1) = \{p\}$. Thus, even though we had only two premises in our Σ (one of which was the negation of the other) it was the one belonging to the more entrenched set that was retained. $\dashv$

Given that we made the assumption that each level (so also Σ_1) is separately consistent, $F(\Sigma_1) = \Sigma_1$.

Example 5. Consider a slightly more complicated example, where Σ is composed of $\langle \Sigma_1, \Sigma_2, \Sigma_3 \rangle$, $\Sigma_1 = \{p, q\}$, $\Sigma_2 = \{\neg p \vee \neg q, r\}$ and $\Sigma_3 = \{\neg r\}$. Then $F(\Sigma_1) = \Sigma_1$. $\Sigma_1 \cup \Sigma_2 = \{p, q, \neg p \vee \neg q, r\}$ has three maximally consistent subsets: $\{p, q, r\}$, $\{p, \neg p \vee \neg q, r\}$, $\{q, \neg p \vee \neg q, r\}$. These have only one common element – r, so $F(\Sigma_1 \cup \Sigma_2) = \{r\}$. $\Sigma_1 \cup \Sigma_2 \cup \Sigma_3 = \{p, q, \neg p \vee \neg q, r, \neg r\}$. It has six maximally consistent subsets: $\{p, q, \neg r\}$, $\{p, q, r\}$, $\{p, \neg p \vee \neg q, \neg r\}$, $\{p, \neg p \vee \neg q, r\}$, $\{q, \neg p \vee \neg q, \neg r\}$, $\{q, \neg p \vee \neg q, r\}$ and they have no common element. So $F(\Sigma_1 \cup \Sigma_2 \cup \Sigma_3) = \emptyset$. Thus, $\Sigma^\star = \{p, q, r\}$. That is, Σ_2 "lost" with Σ_1 when it came to $\neg p \vee \neg q$, but "won" with Σ_3 the "fight" about the value of r. $\dashv$

We are ready to define the operation of non-defeated consequence. We say that A is a non-defeated consequence of Σ iff it classically follows from $\Sigma^\star$:

$$\Sigma \vdash_{\mathbf{ND}} A \text{ iff } \Sigma^\star \vdash_{\mathbf{CL}} A$$

8. Dynamic proofs for non-defeated consequence

Playing around with non-defeated consequence is pretty complex: to find out that something follows from a certain belief base you have to survey all relevant subsets, and for each of them check if it is consistent and if it is not a proper subset of another consistent set, find the common elements of such sets and then verify that the supposed conclusion classically follows from the premise set thus obtained.

ND-consequence relation has been around for some time and wasn't originally provided with a proof theory. (Verhoeven, 2003) developed an adaptive logic in the so-called standard format which captures this consequence operation. She also constructed a slightly more user-friendly direct proof theory capturing this consequence operation. Details lie beyond the scope of this paper – what's important is that I will explain and use the latter in what follows. Thus, adaptive logic comes in because it provides **ND** with a proof theory.[15]

The basic elements of a proof are pretty much like in the simple adaptive logic I have already described, but there are some slight modifications. To start with, for any formula $\phi \in \Sigma_i$, I will sometimes write $^i(\phi)$ instead of ϕ, just to keep track of how entrenched a premise is. Sometimes, instead of writing a whole formula in the dependence column I will just write a line number where it occurs in a proof. The proof system employs two main rules. The first says that, roughly speaking, premises can be introduced based on the assumption that they are not proven to be suspicious (I will explain what it means to be proven to be suspicious later). This is marked by introducing a premise in the line, but also adding it in its own dependence column:

PREM If $\phi \in \Sigma_i$, one may introduce a line consisting of an appropriate line number, ϕ, a dash, PREM and $\{^i\phi\}$.

(Dash is just a place-holder for line numbers which a step relies on. In case of premise introduction, the step does not depend on any other line.)

We need only one more rule – the **U**nconditional **R**ule (RU). It tells us that if something classically follows from the premises we have, given the dependencies, then we can introduce it, making sure that the dependence set "accumulates":

[15] Incidentally, the language of the adaptive logic also has a greater expressive power than prioritized belief bases. For example one can express that a certain formula belongs either to Σ_1 or to Σ_2.

Ru If $\phi_1, \ldots, \phi_k \vdash_{CL} \psi$ and $\phi_1, \ldots, \phi_k$ occur in the proof on conditions $\Delta_1, \cdots, \Delta_k$ respectively, one may add a line consisting of the appropriate line number, ψ, the numbers of the lines in which $\phi_1, \ldots, \phi_k$ are derived, and $\Delta_1 \cup \cdots \cup \Delta_k$.

The rule is called "unconditional" because it doesn't allow for introducing new elements into dependence sets – it only preserves the dependencies.

The marking definition which adequately captures **ND**-consequence requires a few preliminary explanations. First of all, proofs proceed in stages: each application of a rule moves us to the next stage. A set of formulas Δ is shown inconsistent at a given stage of a proof if $\perp$ has been derived on the condition Δ at this stage.

Given that a proof is at a certain stage s, $Minic_s(\Sigma)$ is the set of minimal subsets of Σ shown to be inconsistent at stage s. It is important that we look only at minimal suspicious sets, because we want to localize the anomalies as much as possible.

Since Σ_1 is assumed to be consistent, $Minic_s(\Sigma_1)$ will always be empty. If, on the other hand, $Minic_s(\Sigma_1 \cup \Sigma_2)$ is non-empty, it is the formulas from Σ_2 which are to be blamed, and thus the unreliable formulas (from $\Sigma_1 \cup \Sigma_2$) are just $\bigcup Minic_s(\Sigma_1 \cup \Sigma_2) \cap \Sigma_2$ (recall *Minic*-sets are not sets of formulas, but rather families of inconsistent sets of formulas). In general, given a belief base $\langle \Sigma_1, \Sigma_2, \ldots, \Sigma_n \rangle$, the set of formulas unreliable at stage s, U^n_s is the union of the family of sets

$$\begin{aligned} (\bigcup Minic_s(\Sigma_1 \cup \Sigma_2) \cap \Sigma_2) \cup \\ \cup (\bigcup Minic_s(\Sigma_1 \cup \Sigma_2 \cup \Sigma_3) \cap \Sigma_3) \cup \cdots \cup \\ \cup (\bigcup Minic_s(\Sigma_1 \cup \Sigma_2 \cup \Sigma_3 \cup \cdots \cup \Sigma_n) \cap \Sigma_n) \end{aligned}$$

At last, the marking definition: a line i with condition Θ is marked at a stage s if Θ overlaps with U^n_s. Thus, if we introduce a new line into a proof and want to figure out which lines are marked, we have to go over all the lines introduced so far, identify the minimal subsets known to be inconsistent and "cut off" their weakest elements and all moves in the proof that relied on them. Let's take a look at an example.

Example 6. Say our belief set is composed of three sets: $\Sigma_1 = \{\neg p, p \vee q\}$, $\Sigma_2 = \{\neg q\}$ and $\Sigma_3 = \{r\}$. I'll first give the proof and then provide a commentary.

(1)	$\neg p$	–	Prem	$\{^1(\neg p)\}$	
(2)	$p \vee q$	–	Prem	$\{^1(p \vee q)\}$	
(3)	r	–	Prem	$\{^3(r)\}$	$\curlyvee^{8,9}$
(4)	$\neg p \wedge r$	1, 3	Ru	$\{^1(\neg p), {}^3(r)\}$	$\curlyvee^{8,9}$
(5)	$\neg p$	4	Ru	$\{^1(\neg p), {}^3(r)\}$	$\curlyvee^{8,9}$
(6)	$\neg q$	–	Prem	$\{^2(\neg q)\}$	$\curlyvee^{10}$
(7)	q	2,5	Ru	$\{^1(p \vee q), {}^1(\neg p), {}^3(r)\}$	$\curlyvee^{8,9}$
(8)	$\bot$	6,7	Ru	$\{^1(p \vee q), {}^1(\neg p), {}^3(r), {}^2(\neg q)\}$	$\curlyvee^{8,9,10}$
(9)	q	1,2	Ru	$\{^1(\neg p), {}^1(p \vee q)\}$	
(10)	$\bot$	6,9	Ru	$\{^1(\neg p), {}^1(p \vee q), {}^2(\neg q)\}$	$\curlyvee^{10}$

Up to line (8) the proof develops normally without any lines being marked. After the introduction of line (8), however, the situation changes. $Minic_8(\Sigma_1) = \emptyset$ (by definition), $Minic_8(\Sigma_1 \cup \Sigma_2) = \emptyset$, but $Minic_8(\Sigma_1 \cup \Sigma_2 \cup \Sigma_3)$ is a singleton containing as its only element set $\{^1(p \vee q), {}^1(\neg p), {}^3(r), {}^2(\neg q)\}$ which is identical with $\bigcup Minic_8(\Sigma_1 \cup \Sigma_2 \cup \Sigma_3)$. Clearly, $Minic_8(\Sigma_1 \cup \Sigma_2 \cup \Sigma_3) \cap \Sigma_3 = \{^3(r)\} = U^3_8$ so at stage 8 all lines in whose dependence line $^3(r)$ occurs are marked.

It is quite clear that ultimately $^3(r)$ is not responsible for the contradiction. Once the proof is developed a bit further, up to line (10), the situation changes. $Minic_{10}(\Sigma_1 \cup \Sigma_2 \cup \Sigma_3)$ is identical to $Minic_{10}(\Sigma_1 \cup \Sigma_2)$ and contains only one element: the dependence set from line 10. It has been shown to be inconsistent, and it's a proper subset of the dependence set from line 8 (which was previously shown inconsistent). Thus, $\bigcup Minic_{10}(\Sigma_1 \cup \Sigma_2 \cup \Sigma_3) \cap \Sigma_3 = \emptyset$ and $Minic_{10}(\Sigma_1 \cup \Sigma_2) \cap \Sigma_2 = \{^2(\neg q)\} = U^3_{10}$. Thus, all and only those lines which contain $^2(\neg q)$ are marked in stage 10.$\dashv$

This should suffice as an exposition of what dynamic proofs for the non-defeated consequence operation look like (final derivability is defined on page 12).[16]

This indicates that dynamic proofs are, in a sense, tentative. To really know that something follows from the premises you not only have to derive it in an unmarked proof line, but also to know that the above-mentioned conditions are satisfied. In the propositional case, final derivability is decidable for finite premise sets. In the predicative case (needed further on in this paper) the issue is not in general decidable

[16] (Verhoeven, 2003) proves that the proof theory indeed captures this consequence operation. Also, the way I described the proof theory is not in the *standard format*. Showing that the proof theory can be given a standard-format formulation is beyond our current interests. Details can be found in Verhoeven's paper.

and one needs to reason in meta-language to establish that the above conditions are met.

Thus, for more interesting languages (for which the classical consequence operation is not decidable) dynamic proofs, rather than providing us with ultimate reasons to accept our conclusion, provide us with a systematic method of developing our insight into our belief set.

Some would insist that this fact indicates that what we're dealing with is not logic. This of course hinges on what you mean by logic. In general, discussions whether something counts as logic or not sometimes seem to me rather verbal, so only a few brief remarks will have to suffice.

If what someone cares about is decidability or proof-theoretic manageability of their system, then dynamic logics of the sort discussed above won't be their thing. But it is unclear whether computational considerations are to be decisive here (after all, classical first-order logic is only semi-decidable and second-order logic is not even axiomatizable; yet they are still called logics). Some other important factors indicate at least some degree of logicality of the systems: the consequence operation is well-defined and formal. Particular steps in the proofs are justified by formally described steps and whole proofs are clearly rule-driven (and the rules employed, even though they don't guarantee the truth of the tentative conclusion, certainly help to exclude problematic premises and to increase the reliability of the conclusion as the proof proceeds and insight is gained, in certain cases even informing us that the conclusion is finally derivable). And the fact that just because at a given point we might have to reject something we were led to accept some time before, although causing serious computational difficulties, makes the systems more capable of modeling real human reasoning. The fact is that we often gain our insight into the logical wealth of our premises only gradually and our insight into the logical structure of our beliefs is rarely complete.

9. Dynamic proofs and Galileo's TE

Now that we have described the formal framework, it's time to get back to Galileo's TE and show how the factors mentioned in sections 4 and 5 can, within this framework, be modeled better than by classical means.

Suppose we quantify over bodies freely falling from a certain fixed height in the same external conditions. The language is first-order, it contains two name constants l and s for the large cannon ball and the small musket ball. We have a binary function symbol $+$ which denotes the operation of joining falling bodies (I'll assume that joining x and y results in the same object as joining y and x.) and two unary function symbols W and S; $W(x)$ is the weight of x and $S(x)$ is the speed of x (we assume there are no problems with these denoting functions). Three binary predicates are involved:

- The identity symbol '=',
- The predicate 'D'; '$D(x, y)$' means that there are no differences in material and shape of x and y that would impact the relation between their rate of falling (so, e.g. it is not the case that one is made of cotton wool and the other is made of lead, or that one is really flat and the other round etc.) I will assume that the relation D is symmetrical.[17]
- The relational predicate '$>$'; '$x > y$' means that x is greater than y (we'll be comparing weights and speeds, I allow obvious notational variants).

To shape the premises into a belief base we have to stratify them according to their entrenchment level. This involves certain complications. The main clue as to what the degrees of entrenchment Galileo (or the Aristotelian) assigned to the premises involved is obtained *post factum* by studying his reasoning (and the dropped premises). It is still possible that Galileo felt the need to assign different degrees of entrenchment to the premises only after discovering an inconsistency. Such shifts of entrenchment levels are not modeled in the reconstruction. Also, the study of Galileo's reasoning does not inform us about the priorities of the non-rejected premises. Often we may reach information about hose by carefully studying the context (other writings of Galileo and other writings from the same period). But even in the

[17] (Galileo Galilei, 1638) carefully emphasizes the assumption that the shape of objects is not supposed to be taken under consideration:

> Aristotle declares that bodies of different weights, in the same medium, travel (in so far as their motion depends upon gravity) with speeds which are proportional to their weights; this he illustrates by use of bodies in which it is possible to perceive the pure and unadulterated effect of gravity, eliminating other considerations, for example, figure as being of small importance, influences which are greatly dependent upon the medium which modifies the single effect of gravity alone. Thus we observe that gold, the densest of all substances, when beaten out into a very thin leaf, goes floating through the air; the same thing happens with stone when ground into a very fine powder. But if you wish to maintain the general proposition you will have to show that the same rate of speeds is preserved in the case of all heavy bodies, and that a stone of twenty pounds moves ten times as rapidly as one of two; but I claim that this is false and that, if they fall from a height of fifty or a hundred cubits, they will reach the earth at the same moment. [p. 109]

absence of such information, certain priorities do not matter for the reconstruction.

For these reasons the reconstruction that follows, despite improving on the straightforward accounts, still involves certain simplifications and idealizations. It already takes the entrenchment level of the assumption that was dropped to be fixed in the beginning of the argument. I also impose a certain entrenchment ordering even on those premises which were retained. To some extent, this only mirrors the intuitions I have about the plausibility of the premises involved. I do hope most of the readers will share those intuitions, but (as it should become clear by the end of this paper) nothing essential hinges on any particular ordering of the retained premises: the main philosophical point holds, it's only the particulars of the argument that have to be reconstructed differently.

After these general remarks, let's take a look at the Galilean TE itself. One plausible way to stratify the premises is to divide them into four groups. First, we have the most entrenched beliefs:

- For no two objects x and y the speed of x can be simultaneously smaller and greater than the speed of y.

- Objects considered in the TE either fall with the same speed, or one is falling faster than another.

These are quite entrenched and seem to be conceptually true. Let's write them down as premises of an adaptive proof:

	Σ_1			
(1)	$\forall x, y\, \neg(S(x) > S(y) \wedge S(y) > S(x))$	–	PREM	$\{^1(1)\}$
(2)	$\forall x, y\, (S(x) > S(y) \vee S(x) = S(y) \vee S(y) > S(x))$	–	PREM	$\{^1(2)\}$

Next, we have a few intuitively weaker but still very entrenched beliefs:

- The body obtained by joining two bodies will be heavier than any of those bodies separately.

- The large cannon ball is heavier than the small musket ball.

- There is no relevant difference in shape or material between those balls.

- If there is no relevant difference between two objects, there is no relevant difference between any of these objects and the result of joining them together.[18]

Σ_2

(3)	$\forall x, y\, W(x+y) > W(x)$	-	Prem	$\{^2(3)\}$
(4)	$W(l) > W(s)$	-	Prem	$\{^2(4)\}$
(5)	$\neg D(l, s)$	-	Prem	$\{^2(5)\}$
(6)	$\forall x, y\, (\neg D(x, y) \to \neg D(x+y, x))$	-	Prem	$\{^2(6)\}$

Further, we have two assumptions which are still more entrenched than the Aristotelian assumption, but I find them less compelling than those in Σ_2 (the reader is free to differ and to change this particular detail in the proof, this won't impact the main point).

- The first one says that if there is no difference in material or shape (in the relevant sense) between two bodies, then the lighter one will not fall faster than the heavier one.
- The second one says that if y is faster than x, then joining those objects will result in an object slower than y.

Σ_3

(7)	$\forall x, y\, [\neg D(x, y) \to (W(x) > W(y) \to \neg S(x) < S(y))]$	-	Prem	$\{^3(7)\}$
(8)	$\forall x, y\, (S(x) > S(y) \to S(x+y) < S(x))$	-	Prem	$\{^3(8)\}$

Finally we have (a simplified version of) the Aristotelian assumption: if no relevant shape/material difference occurs, one object falls faster than the other iff it is heavier than the other:

Σ_4

(9)	$\forall x, y\, [\neg D(x, y) \to (W(x) > W(y) \equiv S(x) > S(y))]$	-	Prem	$\{^4(9)\}$

Observe that (9) might *prima facie* be on a par with (7) and (8). What can be said for assigning (9) to a lower level? For one thing, (7)

[18] This is false if the objects do not fall in vacuum. But here we just follow Galileo in his claim (cited in footnote 17) that we are to ignore such factors.

is a weakening of (9). For another, when the Aristotelian is faced with the destructive part, they abandon (9) and not (7) or (8), which (at least *post facto*) shows that if the reconstruction is to be correct, (9) has to be weaker than (7) and (8) after all. (As we will see, once (9) is rejected and (7) and (8) retained, there is nothing mysterious about reaching the positive solution: it simply follows from the premises.)

Let's proceed with the proof now. First, thanks to (4), (5) and (9), we infer that l is moving faster than s.

(10) $S(l) > S(s)$ 4, 5, 9 RU $\{^2(4), ^2(5), ^4(9)\}$

Now, (8) together with (10) entail that joining l with s (which is slower than l) will yield a body that will be slower than l on its own.

(11) $S(l+s) < S(l)$ 8, 10 RU $\{^2(4), ^2(5), ^3(8), ^4(9)\}$

(3) gives us the conclusion that $l+s$ is heavier than l itself.

(12) $W(l+s) > W(l)$ 3 RU $\{^2(3)\}$

(5) and (6) entail that no relevant difference in shape or matter between l and $l+s$ occurs.

(13) $\neg D(l+s, l)$ 5, 6 RU $\{^2(5), ^2(6)\}$

Now, (9) with lines (12) and (13) delivers us:

(14) $S(l+s) > S(l)$ 9, 12, 13 RU $\{^2(3), ^2(5), ^2(6), ^4(9)\}$

which together with (1) contradicts line 11. This shows that a certain set is inconsistent.

(15) $\perp$ 1, 11, 14 RU $\{^1(1), ^2(3), ^2(4), ^2(5), ^2(6), ^3(8), ^4(9)\}$

This, with our marking definition, means we have to cancel our commitment to line 9 and all the inferences that depend on this line.

Moreover, this means that since $\{(1), (3), (4), (5), (6), (8), (9)\}$ is inconsistent, $\{(1), (3), (4), (5), (6), (8)\}$ entails the negation of (9). That is, we not only can cancel our commitment to the Aristotelian assumption, but we also can explicitly reject it, as long as we trust the more entrenched premises involved. Thus, we obtain the following situation:

(1)	$\forall x, y\, \neg(S(x) > S(y) \wedge S(y) > S(x))$	-	PREM	$\{^1(1)\}$	
(2)	$\forall x, y\, (S(x) > S(y) \vee S(x) = S(y) \vee$ $\vee S(y) > S(x))$	-	PREM	$\{^1(2)\}$	
(3)	$\forall x, y\, W(x+y) > W(x)$	-	PREM	$\{^2(3)\}$	
(4)	$W(l) > W(s)$	-	PREM	$\{^2(4)\}$	
(5)	$\neg D(l, s)$	-	PREM	$\{^2(5)\}$	
(6)	$\forall x, y\, (\neg D(x, y) \rightarrow \neg D(x+y, x))$	-	PREM	$\{^2(6)\}$	
(7)	$\forall x, y\, [\neg D(x, y) \rightarrow (W(x) > W(y) \rightarrow$ $\rightarrow \neg S(x) < S(y))]$	-	PREM	$\{^3(7)\}$	
(8)	$\forall x, y\, (S(x) > S(y) \rightarrow S(x+y) < S(x))$	-	PREM	$\{^3(8)\}$	
(9)	$\forall x, y\, [\neg D(x, y) \rightarrow (W(x) > W(y) \equiv$ $\equiv S(x) > S(y))]$	-	PREM	$\{^4(9)\}$	Υ
(10)	$S(l) > S(s)$	4, 5, 9	RU	$\{^2(4), {}^2(5), {}^4(9)\}$	Υ
(11)	$S(l+s) < S(l)$	8, 10	RU	$\{^2(4), {}^2(5), {}^3(8), {}^4(9)\}$	Υ
(12)	$W(l+s) > W(l)$	3	RU	$\{^2(3)\}$	
(13)	$\neg D(l+s, l)$	5, 6	RU	$\{^2(5), {}^2(6)\}$	
(14)	$S(l+s) > S(l)$	9, 12, 13	RU	$\{^2(3), {}^2(5), {}^2(6), {}^4(9)\}$	Υ
(15)	$\perp$	1, 11, 14	RU	$\{^1(1), {}^2(3), {}^2(4), {}^2(5),$ ${}^2(6), {}^3(8), {}^4(9)\}$	Υ
(16)	$\neg(9)$	$1, 3, 4,$ $5, 6, 8$	RU	$\{^1(1), {}^2(3), {}^2(4), {}^2(5),$ ${}^2(6), {}^3(8)\}$	

So far so good. We're done with the destructive part of the TE and we avoided logical explosion: it is not the case that right now we can infer any sentence whatsoever — the contradiction depended crucially on (9), and we retracted our commitment to this premise and all the steps that depended on its truth. However, we can also infer that in fact s and l will be falling at the same rate.

(17)	$\neg S(l) < S(s)$	4, 5, 7	RU	$\{^2(4), {}^2(5), {}^3(7)\}$
(18)	$S(l) > S(s) \rightarrow S(l+s) < S(l)$	8	RU	$\{^3(8)\}$
(19)	$\neg S(l+s) < S(l)$	7, 12, 13	RU	$\{^2(3), {}^2(5), {}^2(6), {}^3(7)\}$
(20)	$\neg S(s) < S(l)$	18, 19	RU	$\{^2(3), {}^2(5), {}^2(6),$ ${}^3(7), {}^3(8)\}$
(21)	$S(s) = S(l)$	2, 17, 20	RU	$\{^1(2), {}^2(3), {}^2(4),$ ${}^2(5), {}^2(6), {}^3(7), {}^3(8)\}$

This completes the proof. Contrary to Brown's claims, we were able to use one and the same structural, rule-driven and formalized argument to derive the antinomy, reject one of the premises, and then, without any (as Norton would have it) stumbles, to reach the desired conclusion.

10. Final remarks

Brown argues against the representability of TEs by means of arguments (the claim that TEs are not arguments arguably follows from this claim: if TEs cannot be sensibly described as arguments, they are not arguments). Brown rests his case on examples of TEs in which a contradiction was encountered and a positive solution was nevertheless reached. Thus, ultimately, the case rests on the claim that making sensible use of a contradiction or rejecting previously held beliefs and coming to an agreement cannot be modeled in terms of a formal logical system. This claim, I suggest, has been shown false: there are plausible formal systems which handle the phenomena Brown refers to in his arguments.

A separate question is whether this has any direct bearing on the "nature of TEs". Is conducting arguments *really* what happens when we use a TE? Assuming this question makes sense and there are methods which in principle would allow us to settle it, our considerations aren't one of them. The goal of this paper was to criticize an argument against the argument view of TEs, not to support the argument view directly. What has been argued for is the *representability* of relevant TEs in terms of arguments, not their *identity* with arguments.

An analogy might help to clarify this point. Consider the discussions surrounding mind-reading. It is rather clear that very often we are able to predict other people's behavior. One view, the theory-theory view suggests that this is because we have in mind a certain (folk) theory of what people do in certain circumstances and use it to make predictions. Its main opponent, the simulationist view, insists that rather than using a theory and propositional reasoning, we just "put ourselves in other person's shoes" (whatever this would consist in) and extrapolate a non-propositional simulation of what we would do in such circumstances. Whatever the outcome of this debate is, this doesn't impact the usefulness of formulating a theory of what people do given certain circumstances to try to predict their behavior. Even if the theory-theory view is false and normally we don't use any theory to predict others' behavior, it is still scientifically interesting whether such a theory can be formulated and how successful (or unsuccessful) it is.

Similarly, the debate about the "real nature" of TEs (however it is to be settled) doesn't have any direct impact on the usefulness of having an explicit theory which helps to make predictions about what the TEs are used to make predictions about. For instance, even if the Galilean TE "really" wasn't an argument, this doesn't mean there is no point in constructing a very closely related argument. Quite to the contrary: just like psychology meant to help us predict human behavior has to be done propositionally, physics has to be done propositionally too.

References

Batens, D. (1995). Blocks. The clue to dynamic aspects of logic. *Logique & Analyse*, 150-152:285–328.

Batens, D. (2004). The need for adaptative logics in epistemology. In Rahman, S., Symons, J., Gabbay, D., and Bendegem, J., editors, *Logic, Epistemology, and the Unity of Science*, pages 459–485. Kluwer.

Batens, D. (2007). A universal logic approach to adaptive logics. *Logica Universalis*, 1:221–242.

Benferhat, S., Dubois, D., and Prade, H. (1997). Some syntactic approaches to the handling of inconsistent knowledge bases: A comparative study part 1: The flat case. *Studia Logica*, 58(1):17–45.

Benferhat, S., Dubois, D., and Prade, H. (1998). Some syntactic approaches to the handling of inconsistent knowledge bases: A comparative study Part 2: The prioritized case. In Orlowska, E., editor, *Logic at work*, volume 24, pages 473–511. Physica-Verlag, Heidelberg.

Brown, J. (1991a). *The Laboratory of the Mind. Thought Experiments in the Natural Science*. Routledge.

Brown, J. (1991b). Thought experiments: A Platonic account. In Horowitz, T. and Massey, G., editors, *Thought Experiments in Science and Philosophy*. Center for Philosophy of Science, University of Pittsburgh, Rowman & Littlefield.

Brown, J. (2004). Why thought experiments transcend experience. In *Contemporary Debates in Philosophy of Science*, pages 23–43. Blackwell.

Brown, J. R. (2008). Thought experiments. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.

Galileo Galilei (1638). *Dialogues Concerning Two New Sciences*. (translation by H. Crew and A. de Salvio published in 1914 by Dover Publications).

Gendler, T. (1998). Galileo and the indispensability of scientific thought experiment. *The British Journal for the Philosophy of Science*, 49:397–424.

Gendler, T. (2004). Thought experiments rethought – and reperceived. *Philosophy of Science*, 71:1152–1163.

Gendler, T. (2007). Philosophical thought experiments, intuitions, and cognitive equilibrium. *Midwest Studies in Philosophy of Science*, 31:68–89.

Herbert, S. (2008). Satisficing. In Durlauf, S. and Blume, L., editors, *The New Palgrave Dictionary of Economics*, pages 243–5.

Kuhn, T. (1977). A function for thought experiments. In *The Essential Tension: Selected Studies in Scientific Tradition and Change*, pages 240–265. University of Chicago Press.

Norton, J. (1996). Are thought experiments just what you thought? *Canadian Journal of Philosophy*, 26:333–366.

Norton, J. (2004a). On thought experiments: Is there more to the argument? *Philosophy of Science*, 71:1139–1151.

Norton, J. (2004b). Why thought experiments do not transcend empiricism. In Hitchcock, C., editor, *Contemporary Debates in the Philosophy of Science*, pages 44–66. Blackwell.

Schrenk, M. (2004). Galileo vs. Aristotle on free falling bodies. *Logical Analysis and History of Philosophy*, 7:1–11.

Verhoeven, L. (2003). Proof theories for some prioritized consequence relations. *Logique et Analyse*, 183-184:325–344.

Acknowledgements

I would like to express my gratitude to Christian Strasser, Frederik Van De Putte, Erik Weber, Rawad Skaff, Dagmar Provijn and Joke Meheus for reading and discussing with me earlier versions of this manuscript. I also owe gratitude to all the people who discussed this topic with me: Graham Priest, Diderik Batens, Anouk Barberousse, Peter Simons, Margherita Arcangeli, Gillman Payette, and the audiences in Geneva, Paris and Ghent, where I gave talks based on this material. I also would like to thank two anonymous referees whose comments were invaluable.