

Why Newton's Laws Are Three

Bryan W. Roberts

A key reason behind Newton's decision to enunciate his first law separately from his second may be a desire to underline causal-talk in his first law in contrast with the first law of Descartes.

It is a commonplace remark in elementary physics texts that Newton's first law of motion is simply a special case of his second law, $F = ma$, when $F = 0$. This is not strictly correct, since the special case $F = 0 = m_0$ is not itself a statement of Newton's first law. What is meant is that the first law seems to be *derivable* from a special case of the second, in a semi-formal sense. The derivation proceeds as follows: suppose we have a body with mass m in some inertial state, and consider the special case of zero force on that mass. Then $a = 0$, by the second law. Finally, we may conclude that zero force implies zero acceleration, or change in motion. This may be alternatively worded as an expression of Newton's first law: a body preserves its state of motion except in the presence of some force.

The claim that Newton's first law is thus derivable from his second law goes back at least to Rouse Ball, who wrote that "it is not clear why [the first law] was enunciated as a separate law" (Rouse Ball 1893, 77). The goal of this paper is to suggest that the primary purpose of the first law, as Newton actually states it, is to enunciate a cause-like notion. Furthermore, this being the case, we will see that the above procedure *cannot* lead to a derivation of the first law from the second. Unfortunately, in recent decades, most attempts at answering Rouse Ball have begun by *accept* that this derivation is possible, and then ascribe various motivations that Newton may have had in giving his first law separately. It has been argued, for example, that Newton may have simply wanted to emphasize the importance of the first law, or that he was using it to provide justification for special cases of his third law (Perl 1966, 585) However, the urgency of these rescue-attempts will evaporate once we have a clear articulation of why Newton's first law cannot in fact be derived from his second.

One attempt at this that is worth reviewing is given by Bernard Cohen (1999). Unfortunately, although Cohen successfully argues that Newton's first law is not an

immediate *special case* of the second, he nevertheless fails to block the *derivation* of the first law as a consequence of the second. Cohen points out that while Newton's first law applies to forces in general, the formulation that he gives of the second law is meant to apply to an *impulsive* force, in which the length of time is vanishingly small. In other words, Newton does not actually say that $F = ma$; he says that $F_i = \Delta mv$. Cohen thus concludes: "law 1 is not a special case of law 2, since law 1 is concerned with a different kind of force" (Cohen 1999, 110).

Although this argument is in principle correct, it nevertheless misses the more interesting claim that is at stake, that the first law may still be *derivable* as a consequence of the second. Let us write down the obvious definition that $F_i = F\Delta t$. Then we may indeed derive $F = ma$ as a direct consequence of Newton's second law, and thus proceed to derive the first law as was done above. This or some similar definition *must* have been implicitly assumed understood by Newton, for without it, it is not possible to deduce the general rule $F = ma$ from Newton's laws. Yet Newton was clearly aware of this more general consequence of his second law, as he makes extensive use of it throughout the *Principia* (see, for example, book 2 prop. 24). Thus, Cohen has not gotten us very far in understanding why Newton needed to enunciate both a first and second law; his insight does not block the derivability of the first law from the second. Of course, one might nevertheless pursue Cohen's line by arguing that Newton was somehow unaware of this derivation, but nevertheless made use of $F = ma$. However, this response is not very compelling, and there is a better explanation available.

Recall the sentence that was shown to be derivable in a semi-formal way from the second law: "a body preserves its state of motion except in the presence of some force." Notably, this is not a very faithful statement of the law that Newton himself articulated; a better representation would be: "a body preserves its state of motion unless some force *compels it to change its state*¹. Newton's use of the Latin verb *cogitur* (from *cogo*, meaning "to compel" or "to urge into action") carries more baggage with it than mere implication. Indeed, what this language seems to be spelling out is a notion very much like causation. Certainly it is not a paradigm example of causal language in Latin, such as

¹I have given a my own very literal translation of Newton's first and second laws in an appendix at the end of this paper, together with the original Latin.

the verb *efficio*, but is nevertheless notably stronger than the language of Newton's second law, which makes use only of proportionality (*proportionalem*).

Whether or not Newton intended *cogitur* in the sense of Aristotelian causation (indeed, he almost certainly did not), the word nevertheless provides Newton with a means of “talking causally” in this first law, in a sense that goes beyond mere logical implication. This being the case, we have good reason indeed to deny that the first law is a consequence of the second. For in order to capture the real meaning of the first law, we would need to derive a kind of *causal implication*, such as: “a body preserves its state of motion unless some force *causes* or *compels* it to change its state.” This sentence does not follow from the logical implication derived above, since a causal implication (famously) does not follow from a logical one. Salmon (1999) famously pointed in numerous ways. For example, a rapid drop in barometric pressure is an event that implies, but does not cause (or compel), an oncoming storm. Similarly, what we have derived from Newton's second law is that an impressed force implies, but need not cause (or compel) a change in inertia. Hence, it seems that Newton's first law is not derivable from his second law, a fact that should not really surprise us anyway.

This interesting little incongruity between Newton's first and second laws then seems to underline *cogitur* as playing a central role in his laws of motion. However, one is immediately led who Newton's target is here. How could one understand the existence of an impressed force to imply—but not cause or compel—a change in inertia? The answer is found in Descartes, who Newton studied extensively, and who seems to have believed that God was the cause of any change in inertial state. As we shall see, Newton's use of *cogitur* is a primary element distinguishing his first law from that of Descartes. We may thus understand its independent enunciation as arising in part from Newton's desire to distinguish this difference in his physics.

Many commentators, including Cohen (1999), have noted that Newton may have imported his first law wholesale from Descartes's first two laws of nature, given in the propositions 37 and 39 of *Principia Philosophae*. But let us more carefully examine Descartes's proposition 37: “each thing, so far as it is simple and indivisible, always continues in the same state, as long as it exists, not at any time changing unless by an

external cause².” Clearly, both Newton and Descartes agree that only phenomena external to (or as Newton says, “impressed” upon) a body can bring about a change in inertia. But notably, Descartes refuses to discuss what these external phenomena might be. In fact, nowhere in the *Principia* does Descartes make explicit mention of the nature of these causes. Not surprisingly, this has given rise to much controversy in the literature as to what Descartes really intends these causes to be. Newton, on the other hand, settles all disputes: every change in inertia is the result of a force. Thus, insofar as the answer to the question of what compels an inertial state to change is central in distinguishing the physics of Newton and Descartes, Newton’s *cogitur* lies at the very center of his physics.

As we have now seen, Newton’s laws are three, and for good reason. One must not be too quick to draw the conclusion that Newton’s first law is superfluous; however, neither must be too eager to this idea too quickly. The derivation of the first law from the second is a viable possibility, until we understand that the first law uniquely contains one crucial aspect of Newton’s physical ontology, namely, the existence of forces as *cogere* or cause-like phenomena. That a body may be “compelled” is something that cannot be derived from Newton’s second law, and the entity that it is compelled by cannot be imported from Descartes.

²A transcription of the original text of this passage, along with my translation, is given in an appendix at the end of this paper.

References

- Cohen, B. I. 1999. "A Guide to Newton's Principia," in *The Principia: Mathematical Principles of Natural Philosophy*. Cohen and Whitman, translators. University of California Press: Berkeley.
- Descartes, R. *Oeuvres de Descartes vol. 8: Principia Philosophiae*. Librairie Philosophique J. Vrin: Paris.
- Newton, I. *The Principia: Mathematical Principles of Natural Philosophy; A New Translation*. Cohen and Whitman, translators. University of California Press: Berkeley.
- Perl, M. R. 1966. "Newton's Justification of the Laws of Motion." *Journal of the History of Ideas*, 27:585-592.
- Rouse Ball, W. W. 1893. *An Essay on Newton's "Principia"*. Macmillan and co: New York.
- Salmon, W. 1999. "Scientific Explanation," in *Introduction to the Philosophy of Science*, pp. 22. Hackett: Indianapolis.

Appendix:

Translations of Relevant Passages

In trying to understand the specific terminology of Newton and of Descartes, the present author found it useful to produce translations of the following passages, with the sole intent of remaining as faithful and literal to the original Latin as possible.

Descartes: *Principia Philosophae*, II.37

Prima lex naturae: quod unaquaeque res, quantum in se est, semper in eodem statu perseveret; sicque quod semel movetur, semper moveri pergat.

Atque ex hac eadem immutabilitate Dei, regulae quaedam sive leges naturae cognosci possunt, quae sunt causae secundariae ac particulares diversorum motuum, quos in singulis corporibus advertimus. Harum prima est, unamquamque rem, quatenus est simplex et indivisa, manere, quantum in se est, in eodem semper statu, nec unquam mutari nisi a causis externis. ...

First law of nature: that each thing as long as it exists, always perseveres in the same state; and thus once it is moved, it always continues to move.

Moreover, from this same immutability of God, certain rules or laws of nature can be apprehended, which are secondary and also particular causes of the various motions that we notice in individual bodies. The first of these is that each thing, so far as it is simple and indivisible, always continues in the same state, as long as it exists, not at any time changing unless by an external cause. ...

Newton: *Principia*, Axiom 1

Corpus omne perseverare in statu suo quiescendi vel movendi uniformiter in directum, nisi quatenus illud a viribus impressis cogitur statum suum mutare.

Projectilia perseverant in motibus suis, nisi quatenus a resistentia aeris retardantur, et vi gravitatis impelluntur deorsum. Trochus, cujus partes cohaerendo perpetuo retrahunt sese a motibus rectilineis, non cessat rotary, nisi quatenus ab aere retardatur. ...

Every body perseveres in its state of resting or of moving with uniform direction, unless it is compelled by an impressed force to change its state.

Projectiles persevere in their motion, except so far as they are slowed by the resistance of air, and are thrust downward by the force of gravity. A hoop, having cohesive parts, perpetually draw each other back by rectilinear motion, rotation does not cease, except so far as it is slowed by a i r

Newton: *Principia*, Axiom 2

Mutationem motus proportionalem esse vi motrici impressa, et fieri secundum lineam rectam qua vis illa imprimitur.

Si vis aliqua motum quemuis generet; dupla duplum, tripla triplum generabit, sive simul et semel, sive gradatim et successive impressa fuerit. ...

Motion is a proportional change with impressed force, and occurs along the straight line on which the force is impressed on it.

If some force generates any motion, double will generate double and triple will generate triple, that is all at once, or successively by degrees