Forms And Causes In Plato’s Phaedo

Christopher Byrne

Plato has been accused of many things. With regard to the Phaedo, he has been accused of not knowing what a cause is, or at least of not being able to distinguish between the different types of aitia or causes. Like many of the criticisms of Plato’s doctrines that do not originate with Plato himself, this view goes back to Aristotle. In his Metaphysics and De generatione et corruptione, Aristotle criticizes “Socrates in the Phaedo” for making the Forms into the causes of not only the being of sensible substances, but also their generation and destruction. This, Aristotle says, is impossible because it leaves unexplained both why those things which have no Forms are generated and destroyed, and why those things which do have Forms are only intermittently generated and destroyed.

Some scholars have come to Plato’s defence on this point, most notably Gregory Vlastos in an article entitled “Reasons and Causes in the Phaedo.” Stated quite simply, Vlastos’s defence of Plato is that the Forms are not intended to be causes at all, but only reasons. In other words, Plato never claimed that the Forms act as causes in the sense of the causes of motion and change; rather, they function only as explanatory reasons, in the sense of the principles according to which individual objects are classified as belonging to particular kinds or classes of objects. Stated in Aristotelian language, Plato’s Forms in the Phaedo are intended to act only as formal causes, not as final, efficient, or material causes. Thus, Plato is not guilty of confusing reasons with causes as some of his critics, beginning with Aristotle, have charged.1


2. De generatione et corruptione II 9, 335b7-16; Meta. I 9, 99b3-9; XII 5, 1079b14-15 & 1080a2-8; cf. also Meta. I 9, 99a24-26; VIII 6, 1045b7-9; XII 6, 1071b14-17.


The advantage of Vlastos’s view is that it makes Plato look far more plausible on this question. Its disadvantage is that it leaves somewhat unclear why anyone thought in the first place that Plato was talking about causality in the Phaedo, except perhaps because people were misled by the ambiguity of the Greek word aitia. One would have thought that at least Aristotle, who was so careful to distinguish between the various meanings of this term, would have been sensitive to its ambiguity when discussing the Phaedo, and if not Aristotle, then at least those who have read Aristotle. Indeed, on Vlastos’s account, Aristotle’s reading of the Phaedo is especially perverse because Aristotle is the authority to whom Vlastos appeals to show that the Greek word aitia means more than just an efficient cause.5
is one in which the causal relations between sensible substances simply reproduce the logical relations between the Forms. As a result, Vlastos argues, Plato reduces the notion of causal necessity to that of logical necessity.\footnote{Evan L. Burge argues in a similar fashion, and, like Vlastos, claims that Plato ignores the difference between logical and physical necessity, and "assimilates all necessity to logical necessity" (ibid., p. 8).}

In the following, I attempt to do two things. The first is to show that, although Vlastos is right to think that the Forms by themselves are not causes in the sense of the causes of motion and change, he has, nevertheless, misunderstood the role of the Forms in the explanation of motion and change as described in the *Phaedo*. In particular, the explanation of motion and change is not simply a matter of logical necessity, as Vlastos would have it. I have chosen to include this critical discussion of Vlastos's views because he arrives at the above conclusion on the basis of an interpretation of the *logoi* or definitions of the Forms which is widespread in the literature, an interpretation which might be broadly characterised as neo-Kantian.\footnote{The classic statement of this position is found in Paul Natorp's *Plato's Identitäten* (2nd ed. 1922); reprt. Darmstadt: Wissenschaftliche Buchgesellschaft, 1975), esp. pp. 133-36, 157-61; cf. also Joseph Moreau, *La construction de l'idéalisme platonicien* (Paris, 1939; reprt. Hildesheim: Georg Olms, 1967), esp. pp. 309-18, 381-89. A critical response to this position is presented by Nicolai Hartmann, "Das Problem des Apriorismus in der Platonischen Philosophie," *Sitzungsberichte der Preussischen Akademie der Wissenschaften*, XV (1935), reprt. in his *Kleinere Schriften*, Vol. II (Berlin: de Gruyter, 1957).}

The second task is to offer an alternative account of the relation between the Forms and the explanation of change, one which shows both how the *Phaedo* account anticipates important features of Aristotle's own account of change and also how Aristotle's remarks about the *Phaedo* are not as preposterous as they first appear.

When one turns to the text of the *Phaedo*, there are excellent reasons for thinking that Plato here was discussing efficient causality and the relation between Forms and efficient causes. To begin with, the context in the *Phaedo* in which Socrates discusses the reasons for his turning to the Forms does have an intimate connection to the whole question of efficient causality. Socrates indicates at the beginning of this section of the dialogue that they have to consider the whole question of the cause of generation and corruption (95e9). He then indicates that, in turning to natural science or the "inquiry into nature," he sought to learn why each thing comes to be, perishes and is (96a8-10). On Vlastos's reading, in later turning away from pre-Socratic natural science and turning toward the Forms, Socrates abandoned his concern with the causes of generation and corruption, and devoted his attention only to the question of what makes things to be the kind of things they are. In other words, Socrates lost interest in the efficient and teleological causes that would be necessary to explain the universe in the way in which he originally thought Anaxagoras did.

If this is in fact what happens, then, within the context of the dialogue, the discussion of the Forms must be judged a complete failure. For this section of the dialogue is intended to be a response to an argument put forth by Cebes against the immortality of the soul. Cebes's argument is stated in the form of an analogy: the relation between the soul and the body is the same as that between a weaver and a cloak woven by him. Just as the weaver turns yarn into a cloak by weaving and shaping it, so too the soul shapes and moulds the parts of the body into a certain arrangement. The point here is not that the weaver is the shape or form of the cloak, but rather that the weaver gives to the cloak its distinctive structure and form. What Cebes is getting at here is the role of the soul as that which moves, directs and controls the body.

Heretofore in the dialogue Socrates has been able to avoid this question. The previous arguments for the immortality of the soul stressed the separation of the soul from the body. The persuasive force of these arguments was based on the claim that the soul is capable of exercising its distinctive activity of knowing apart from the body, and thus is capable of existing apart from the body. Cebes's objection points out that this emphasis on the cognitive activity of the soul leaves unanswered how the soul exercises its rule and direction of the body. Socrates has already committed himself to the view that the soul does exercise such a function; it is central to his second argument against Simias's view that the soul is the harmony of the body (94b-e). For if the soul is nothing but the proper arrangement of the parts of the body, it would be impossible for the soul to direct and control the body, something all participants in the discussion agree the soul does. The difficulty raised by this, however, is how the soul can rule over the body without being in some sense connected to it, or at least in contact with it. It is this contact with the body which Cebes thinks will lead to the destruction of the soul. He describes it as being worn out and dying after it enters the body (88a8-10).

What Socrates has to show is how the soul can be both something like an efficient cause, moving and directing the body, but also something immortal which is not corrupted by this rule over the body. If he does not do this, Cebes's argument stands unrebuted.

Socrates's response to Cebes can be divided into roughly three parts: 1) the aporetic part, describing his early interest in natu-
ral philosophy and subsequent disappointment with his predecessors; 2) Socrates’s turn away from examining things through the bodily senses and the introduction of the “safe” hypothesis of the Forms; 3) the “cleverer” or more sophisticated answer to “why”-questions. The third part is the focus of the exegetical dispute because it is here, if anywhere, that Socrates introduces a connection between Forms and causes. But the ground for this is prepared in the two earlier parts, which we shall consider in the order set out above.

With respect to the first, aporetic part, the most important question for our purpose is what lesson Socrates draws from his consideration of the views of his philosophical predecessors, a lesson which in turn guides his own method of inquiry or “second best journey.” Central to Vlastos’s argument is his claim that this encounter with pre-Socratic natural philosophy leads Socrates to abandon teleology as the basis for the inquiry into the causes of motion and rest in nature. But there are several problems with this interpretation. First of all, it is based on a mistranslation of a key passage in the text: At 99c-d, Vlastos has Socrates say, “since I have been denied this atitia and have failed to either find it myself or learn it from another” (emphasis added), whereas in fact Socrates says: “I was denied this atitia and failed to either find it myself or learn it from another.” In other words, Vlastos translates this passage as if it were in the perfect tense and Socrates were speaking from his present point of view, whereas the two verbs at 99c8-9 are both in the aorist tense and Socrates is talking about something that happened in the past. This leaves open the possibility that Socrates’s “second best journey” “does in fact lead him in the direction that he had originally expected from the earlier natural philosophers. But this Vlastos resolutely rejects. Vlastos sees Socrates’s new method of inquiry, his “second best journey,” as “an alternative method of searching for atitia” rather than as “an alternative method of searching for teleological atitia.” Vlastos is certainly correct when he says that we should not assume that for Plato teleological causes are “the only admissible atitia of anything whatever.” But we are not talking about the causes of “anything whatever”; as Vlastos himself indicates, teleological causes are for Plato “the preferred (most fundamental, most illuminating) explanations of natural phenomena,” and the natural phenomena in question presumably include that particular natural phenomenon Socrates is attempting to explain in this passage in the Phaedo, namely, the animation, rule, and direction of the body by the soul and the compatibility of this with the immortality of the soul. If, as is indicated by his explanation of why he is in jail waiting to be executed, Socrates is still concerned with teleological causes, then he is presumably also still concerned with the efficient causes directed toward final causes, and not just with formal causes.

It is certainly correct that Socrates’s alternative method is, in the first instance, a search for other causes besides teleological ones, a search which ultimately leads to the hypothesis of the Forms. What Vlastos fails to appreciate is the connection between Forms and teleological causes. He is right to point out that there is little mention of teleology in Socrates’s own account of the causes of things. It may be that in the Phaedo Socrates never resolves the question of the place of final causes in the material universe as a whole. This does not prove that he has abandoned entirely the programme of a teleological science of nature in general, let alone the use of final causes in the explanation of human behaviour. The inadequacy of the previous tradition in not allowing for the Good as a cause seems just as incorrect to him at the end of his life as when he was a young man. Hence, this requirement presumably applies just as much to the method of his “second best journey”, the turn to the Forms. In other words, the turn to the Forms need not be interpreted, as it is by Vlastos, as a turn away from final causes. On the contrary, what the “second best journey” seems to recognize is that the role of the Good and the Fitting in the explanation of nature can be seen only after one has turned to the Forms and their definitions. The Forms of health, beauty and strength mentioned earlier in the dialogue are examples of this connection between knowledge of the Forms and knowledge of the Good since they all involve the notion of the proper functioning and arrangement of the parts of the human body. Thus, all three involve the Form of the Good. The hypothesis of the Forms is not identical to the positing of teleological causes, but it is a necessary condition for the latter.

Another example of Socrates’s continued interest in final causes is found at the beginning of the myth which closes the dialogue. Here Socrates suggests that the earth is at rest at the centre of the universe; the reasons given are the inner “equipoise” or balance of the earth and the homogeneity of the heavens surrounding it. Anything so constituted and situated, he argues, will remain at rest because it has no greater tendency to move in one direction rather than another (109a2-6). This is presented as contradicting the accounts of those pre-Socratics who required the presence of other material bodies acting on the earth to keep it in place. Part of Socrates’s criticism of pre-Socratics such as Anaxagoras was their failure to provide a sufficient cause for the ordered pattern of change that we observe in the material universe. This

can be restated as the principle that for every observed change there has to be an adequate cause. Socrates' explanation of the motionlessness of the earth represents an application of this principle inasmuch as it follows from the latter that in the absence of any such moving cause, there will be no change. It may be, then, that Socrates never did fully satisfy the original expectation he had when he first heard about Anaxagoras' book. But, as the above examples indicate, he does not seem to have abandoned this project and, as I endeavour to show below, the theory of the Forms is an important first step in carrying it out.

With respect to the introduction of the "safe" hypothesis of the Forms, it is to be noted that this method does not apply just to the inquiry into the causes of things, but to any type of inquiry whatsoever (100a6). This is because the purpose of the method Socrates introduces is to thwart those "enemies of reason" (antilogikoi) who are principally interested in generating contradictions rather than in discovering the truth of anything, a remark which seems to be aimed at the Sophists (101d5-e6). The general way in which the sophistical form of argument is overcome consists in, first, distinguishing between the premises and conclusions of an argument, second, not admitting as a premiss a proposition which is either explicitly or implicitly self-contradictory, and third, not admitting as a premiss any proposition which contradicts any of the previously-accepted premisses or anything validly derived from them. As a way of avoiding contradictions, this seems only reasonable and proper, but by itself hardly seems to require anything as elaborate as a theory of Forms. Plato surely is not arguing that only those who subscribe to the theory of Forms can understand deductive inference and the difference between arguing from and arguing to certain first principles.

In contradistinction to the general method of hypothesis outlined above, the hypothesis of the Forms is intended to prevent a particular way of falling foul of the above formal restrictions and thereby generating contradictions. The theory of Forms insists first and foremost upon the distinction between the Forms themselves and the things that participate in the Forms. If one does not make this distinction, it is quite easy to generate contradictions, and Socrates gives us several examples of this. Addition and division, in the sense of combination and separation, are opposed processes, yet something can be made to be two in number by virtue of both of them. Something might be said to be beautiful by virtue of a certain shape or colour, and yet that same shape or colour might make something else appear to be ugly. The generation of these apparent contradictions is particularly easy using relational properties. One person is taller than another by a head and shorter than someone else by a head, and so both tall and short; what is apparently worse is that this person might be said to be tall by virtue of something short, namely a head (101a5-b2). By distinguishing between the relevant Form and the thing that participates in the Form and, specifically, by insisting that something has a particular property only because, and to the extent that, it participates in the corresponding Form, it becomes possible to begin clearing up the various equivocations that lie behind these apparent contradictions.11

By distinguishing between the Form and the thing participating in the Form, one comes to see apparently simple objects as composites. This involves distinguishing between those characteristics of an individual object by virtue of which it participates in a particular Form, on the one hand, and the rest of the object and its properties, on the other. This allows one to avoid not only the kinds of contradictions discussed above, but also other apparent contradictions that might arise in the understanding of motion and change. Plato points to this in the brief exchange Socrates has with a nameless interlocutor which separates his "safe" hypothesis about the Forms from his more sophisticated answer (103a4-c2). Socrates has just claimed that not only will the Forms not admit their opposites, but also the Forms in us will never admit their opposites in us. For example, not only will Tallness never become Shortness, but also the tallness in us will never become shortness in us. At this point, Socrates is accused of contradicting himself; earlier he had said that all change involves something coming to be out of its opposite. Socrates responds by again emphasizing the distinction between the Form and the thing that has, or participates in, the Form. All instances of motion and change involve some one thing participating in contrary Forms at different times. The point here is that unless one gets clear about this distinction between the Forms and the things that participate in the Forms, the nature of change and motion becomes unintelligible.


11. Perhaps the most important equivocation here is the one as to which Forms are really involved; this allows one to pretend that two properties are contradictory when in fact they are not (e.g. one person is both tall and short by virtue of being taller than one person and shorter than another).
In presenting his argument, then, Socrates leads us to distinguish between three different questions: 1) What is it for something to be an x? 2) What is it for something to become an x? 3) What is the cause of something becoming an x? Socrates’s “safe” hypothesis of the Forms provides us with both an answer to the first question and, at the same time, a way of avoiding the specious contradictions of the Sophists. The answer to the second question presupposes the answer to the first question; in order to be able to say what it is for something to become an x, one first has to know what it is for something to be an x. Thus, the answer to the second question requires both the hypothesis of the Forms and the view that in every change, opposite Forms are found successively in one persisting subject of change; the latter is set out in Socrates’s first argument for the immortality of the soul (69e-72e). The answer to the third question, about the causes of change, presupposes an answer to the first two; if by a cause we mean primarily something that produces a change, to understand how change is produced, we first have to know what it is for something to undergo change. The exegetical question we are dealing with here can, then, be formulated more precisely: Does Socrates in his second, “cleverer” answer go on to consider our third question, about the causes of change, and if so, what is the nature of the connection between the Forms and the causes of change?

The first indication that the Forms are involved in efficient causality is found in Socrates’s exchange with his nameless interlocutor mentioned above. When Socrates says that the Forms found in individual substances will not submit to becoming their opposite, but will either be destroyed or retreat, he is not just claiming that one and the same thing cannot have contrary properties. He also seems to be suggesting that the presence of one property can be responsible for the destruction or disappearance of a contrary property in a particular individual. One must grant that the Form of heat never heats anything and, consequently, could never destroy the Form of cold, ignoring for the moment that, as a Form, the latter is already indestructible. But it does seem possible for the heat and cold in individual things to affect and possibly destroy each other.

Nevertheless, if heat and cold are to exercise anything like efficient causality, it can only be because they are found in substances in the sense of particular concrete things; only such substances can bring about changes in other substances, and hence Forms, which are not substances in this sense, are never by themselves enough to be efficient causes. Thus, as several commentators have pointed out, the question becomes whether Socrates is just discussing the relations between Forms here, or the relations between substances as well.  

Vlastos is prepared to admit that what Socrates says here does have implications for the causal relations between substances. But he makes these causal relations to be ones of logical necessity, simply following from the relations of logical entailment holding between the different Forms. In one way, this claim is innocuous. What else does it mean to say that two properties are opposites than to say that they cannot both be found in one thing at the same time. Whatever is hot cannot be cold; whatever is alive cannot be dead. From the notion of two properties as opposites, it is only a short step to the notion that the coming to be of one property in something is also the cause of the destruction of its opposite in that same thing. If things are what they are only to the extent that they participate in various Forms, then it is hardly surprising that the way in which they change and act on one another should reflect the relations between the Forms. This is only to say that causation is grounded in the nature of the things involved.

The reduction of causal necessity to logical necessity comes about only if one thinks of the logoi expressing the nature of the Forms as being themselves logical truths. Given what heat and cold are, it necessarily follows that whatever is hot, is not cold, but the deductive necessity of a valid argument does not turn the conclusion into a logical truth. Only if the logoi themselves are interpreted as logical truths is it the case that all propositions derived from them are logical truths. Specifically, the laws of nature which are entailed by the definitions of the Forms are logical truths only if the original definitions of the Forms are as well. It is here that the neo-Kantian background of Vlastos’s interpretation becomes most apparent: since the definitions of the Forms are not empirical truths on this account, but are discovered a priori, they must be logical truths. These definitions serve not only to specify the necessary and sufficient conditions which an individual object must meet if it is to be classified as belonging to a certain natural kind; they are also a priori truths on a par with the truths of mathematics and logic, and they ground the intelligibility of all sense experience.

A comprehensive examination of this interpretation lies beyond the scope of this paper. For the moment, suffice it to say that it is unclear how, on this account, the Forms can also be thought of as standards or models imitated by less perfect particular objects and actions. For our purpose, it is important to note that Vlastos

ignores certain fundamental ontological differences between the various Forms which bear on the question of efficient causality. It is certainly true, for example, that on Plato's account whatever participates in the Form of fire must also participate in the Form of heat. It is also clear that the relation between fire and heat is not one of species to genus; fire is not a species of heat nor heat a species of fire. Furthermore, a species does not participate in a genus. But fire does participate in the Form of heat, and on the basis of what Socrates has said so far in the dialogue, things that participate in the Forms are different in kind from the Forms themselves. This indicates that the Form of fire and the Form of heat are forms of things which are ontologically different in kind. Further evidence of this is given by Socrates' claim that fire is not the contrary of snow even though they are essentially qualified by contrary properties. This lack of contrariety between Forms such as the Form of snow and the Form of fire suggests that they are Forms of substances as opposed to Forms of attributes, concrete general terms, not abstract ones. This suggests, then, that when Socrates is speaking about the relations between fire and snow, he means particular substances and not just the logical relations between the Forms of certain attributes.

A further difficulty with Vlastos' view, as David Gallop has pointed out in his commentary on the Phaedo, is that it does not work for the relation between the soul and the body, and if it does not properly capture the relation between the body and the soul, then it also does not capture the relation between, say, fire and the body the fire comes to be present in. It might be that when Socrates says a body will be hot when fire comes to be in it, he just means that anything participating in the Form of fire a fortiori participates in the Form of heat. This simply says that all fire must be hot. But Socrates goes on to use the same locution to talk about the soul and body: It is by virtue of the soul coming to be in a body that that body has life. Here Socrates cannot just be saying that whatever participates in the Form of the soul also participates in the Form of life because a body does not participate in the Form of the soul. If it did, it would be a soul. Rather, Socrates seems to be saying that a soul, which is itself already alive, animates or brings to life an otherwise inanimate body. Following this same pattern, when Socrates speaks about a body being hot by virtue of fire coming to be in it, he does not mean that the body comes to participate in the Form of fire, i.e., it becomes fire, but rather that it is the presence of the substance fire in a body that makes what would otherwise be a cold body to be a hot one. In other words, Socrates is talking about how one substance can cause another substance to acquire a certain property and not about the relation between a substance and one or more Forms.

One might object that there is no talk of substances here and, therefore, no ground for talking about the efficient causality of one substance acting upon another. But even if we remain just with Platonic Forms, it is clear that there is an important difference between the way things participate in the Forms of, say, fire, snow and the soul, on the one hand, and the Forms of heat, cold and life, on the other. Whatever participates in the Form of fire is or becomes fire; whatever participates in the Form of heat has heat or is hot, but we would not say that it is heat. Similarly, whatever participates in the Form of life has life or is animated; again, we would not say that it is life. In other words, we have another argument to suggest that the Forms of fire, snow and the soul are the Forms of substances, whereas the Forms of hot, cold and life are the Forms of properties of substances.

When Socrates is talking about fire making a body hot or the soul making a body alive, he is talking, then, in the first instance, about the causal relations between substances and not the logical relations between Forms. This is confirmed by the mathematical example he gives. According to his "cleverer" answer, he will no longer say that a number comes to be odd by participating in the Form of oddness, but by virtue of a unit or monad being present in it. If by a unit Socrates means the Form of oneness or unity, this statement makes no sense. For what it would mean for all the odd numbers to participate in the Form of oneness would be whether all of them would be reduced to the number one, in the sense that each would become a single unit, or that each of the numbers would be something like a unified manifold. But both of these alternatives are unacceptable: in the first case, there would no longer be a manifold of different odd numbers, but just one, namely the number one; in the second case, there would be no difference between odd and even numbers because all numbers already participate in the Form of oneness to the extent that they are unified manifolds, whereas the unit or monad Socrates is thinking about belongs only to odd numbers. The unit or monad in question, then, must be the unit present in all odd numbers which remains when they have been divided by two. As such, it is not a Form in which odd numbers participate.

If our analysis is correct, then when Socrates is discussing the relation between fire and heat, snow and cold, disease and illness, and finally the soul and life, he is not, contrary to what Vlastos says, just discussing the relations between Forms. It is certainly true to say that because, for example, heat is a defining property of

fire, whatever participates in the Form of fire must also participate in the Form of heat. But, if the final argument for the immortality of the soul is to answer Cebes’s final objection, it must be based on the notion that the soul in some way moves and directs the body; this indicates that the relation of the body and the soul is understood here as being one between two substances. Vlastos offers no account of how this is possible because he considers only those causal relations that are grounded in relations of logical entailment between Forms. His account also overlooks the fact that things participate in Forms such as those of fire, snow and the soul in a very different way than they participate in Forms such as heat, cold and life. Once the relation between the soul and the body is seen to be one between two substances, it becomes possible to interpret Socrates’s “cleverer” answer as dealing not only with formal, but also with efficient causality.

This would also explain why Aristotle finds Plato’s account of this topic in the Phaedo to be inadequate. For Plato offers no account there of how these Forms are brought into being in particular substances in the first place; in other words, there is no account of how particular substances exercise their capacity for efficient causality so as to produce changes in one another. In the absence of such an account, Aristotle’s remark that the Forms themselves seem to act as efficient causes makes more sense. Aristotle’s own answer to this question requires an elaborate account of the various ways in which material substances interact and thereby pass along a Form from one substance to another. But before he discusses these questions, Aristotle too has to consider the question of what makes any kind of change possible, and his answer in the Physics in terms of a substratum and a pair of contrary properties is remarkably similar to what Socrates says in the Phaedo. More importantly, Aristotle follows the same sequence of questions as does Socrates, discussing first what it is to be something of a certain kind before discussing what it is to become that kind of thing and finally what causes this to happen. Even if the programme for a teleological science of nature is not carried out in the Phaedo, it does at least provide a guide for what questions are to be asked, what kinds of causes are to be sought and what role the Forms will play in such a science. Aristotle’s own natural science follows this guide.

St. Francis Xavier University
Antigonish, N.S.