# G.E. Moore's Naturalistic Fallacy

by Alonzo Fyfe (20170104)[1]

It will often not take long in many discussions about ethics where one or more participants knows a little about philosophy until one person accuses the other of committing "the naturalistic fallacy". Usually, when it is used, it is immediately followed by the sound of the person who uses it closing his mind – as the person who uses it is confident that they have delivered a fatal blow.

People who use this argument seldom have a clear idea of what "the naturalistic fallacy" is. Oliver Curry identified eight different arguments given the name.[2] I was among those unclear about the fallacy, until I decided I wanted to know.

Here, I am interested specifically in G.E. Moore's naturalistic fallacy – in the mistake he thought he saw in most of the writings of other moral philosophers.

## The Naturalistic Fallacy

Common resources that one may appeal to in trying to learn about the naturalistic fallacy offer answers that I judge to be misleading.

Wikipedia introduces the Naturalistic Fallacy by saying:

> *It would be fallacious to explain that which is good reductively in terms of natural properties such as "pleasant" or "desirable".[3]*

The Stanford Encyclopedia of Philosophy states:

> *At the core of these criticisms is the thesis that the position involves a fallacy, the 'naturalistic fallacy', of supposing that goodness, which Moore takes to be the fundamental ethical value, can be defined in naturalistic terms, in terms, say, of pleasure or desire or the course of evolution.[4]*

Moore does believe that this identifying goodness with a natural property such as pleasure or desire or the course of evolution is a mistake – and he believes that when people attempt to prove moral naturalism they are often guilty of committing the naturalistic fallacy. However, this confuses the fallacy with the mistaken conclusion. This is like saying, where somebody responded that opponent to the thesis of human-caused climate change engaged in *cherry picking* their data, that the fallacy of *cherry picking* is defined as the denial of human-caused climate change. Those attempting to equate goodness from natural properties often commit the naturalistic fallacy, but the naturalistic fallacy does not consist in equating goodness to a natural property.

---

[2] Curry, Oliver (2006), "Who's Afraid of the Naturalistic Fallacy?" Evolutionary Psychology, human-nature.com/ep – 2006. 4: 234-247, http://evp.sagepub.com/content/4/1/147470490600400120.full.pdf, retrieved January 3, 2017.
[3] Wikipedia, "Naturalistic Fallacy", https://en.wikipedia.org/wiki/Naturalistic_fallacy, retrieved 12/26/2016.
[4] Stanford Encyclopedia of Philosophy, "George Edward Moore", https://plato.stanford.edu/entries/moore/#3, retrieved 12/26/2016.

In fact, equating goodness with a natural property is not even a fallacy. A fallacy is a mistaken inference. A person commits a fallacy when that person asserts one or more premises, and a conclusion, and falsely claims that the premises support the conclusion. The *ad hominem* fallacy asserts that because Jim is an alcoholic who was once convicted of embezzlement that we can dismiss his claims about the effects of raising the minimum wage. It does not follow from the premises that a person is an alcoholic and convicted embezzler that his claims concerning the minimum wage are flawed.

## The Naturalistic Fallacy

The naturalistic fallacy, in fact, is this:

The naturalistic fallacy involves a false inference whereby the person making an argument reports as her premises that property A is often (always) found in company of property B, that the claim that something has property A means the same thing as the claim that it has property B.

Let us assume that "being red" and "being sweet tasting" are always found together – as they are with apples and strawberries. An instance of the naturalistic fallacy would be to argue that, whenever we discover that something is red, we also discover that it is sweet tasting. Therefore, the property of "being red" is the same property as the property of "being sweet tasting".

In Moore's own words:

> *Consider yellow, for example. We may try to define it, by describing its physical equivalent; we may state what kind of light-vibrations must stimulate the normal eye, in order that we may perceive it. But a moment's reflection is sufficient to shew that those light-vibrations are not themselves what we mean by yellow. They are not what we perceive. Indeed, we should never have been able to discover their existence, unless we had first been struck by the patent difference of quality between the different colours. The most we can be entitled to say of those vibrations is that they are what corresponds in space to the yellow which we actually perceive.[5]*

We can easily see that this is fallacious reasoning. Two things can occur together and still be two different things.

It may be the case that Moore wants to limit the term "naturalistic fallacy" to instances where this mistake is made specifically involving the term "good". When the mistake occurs with respect to other terms – like "yellow" – it remains a fallacy but would then get another name, or no name at all. Nonetheless, this is the type of invalid inference that Moore wants to call the "naturalistic fallacy".

The way this works out in ethics (according to Moore) is that a moral philosophers notes that goodness is found in pleasure or that which is desired, and asserts from this that goodness is pleasure or that which is desired – that the terms and phrases mean the same thing. A philosopher who does this commits the naturalistic fallacy.

As further evidence that this is the fallacy that Moore calls the naturalistic fallacy – and that the fallacy does not involve equating 'good' to some natural property – note that he says that 'metaphysical ethics'

---

[5] Moore, G.E., *Principia Ethica*, Chapter 1, Section 10, http://fair-use.org/g-e-moore/principia-ethica/chapter-i, Retrieved 12/25/2016.

also commits the naturalistic fallacy. By 'metaphysical ethics', Moore specifically meant that ethics grounded on a belief that natural properties are not the only properties that exist.

> *What, then, is to be understood by 'metaphysical'? I use the term, as I explained in Chapter II, in opposition to 'natural.' I call those philosophers preeminently 'metaphysical' who have recognised most clearly that not everything which* is *is a 'natural object.*[6]

However, metaphysical ethicists still commit the naturalistic fallacy:

> *[Metaphysical ethicists] also imply, as I said, that this ethical proposition follows from some proposition which is metaphysical: that the question 'What is real?' has some logical bearing upon the question 'What is good?' It was for this reason that I described 'Metaphysical Ethics' . . . as based upon the naturalistic fallacy.*[7]

This is, in fact, a fallacy. One cannot prove, from the fact that A always shows up with B, that being A is the same thing as being B.

## The Open Question Argument

This interpretation of the naturalistic fallacy also makes sense of Moore's 'open question argument' – which he offers as a test for the naturalistic fallacy.

Recall that a philosopher commits the naturalistic fallacy when he takes the co-occurrence of one property (e.g., goodness) with another (e.g., pleasure) as reason to believe that they are two terms for the same thing. We can determine that they are not, in fact, two terms for the same thing by the fact that the question, "X is A, but is X also B?" is an 'open question'. It is a question that makes sense.

For instance, we can distinguish "being red" from "being sweet-tasting" because the question, "That is red, but is it sweet tasting?" is an open question. Thus 'red' cannot be the same thing as 'sweet-tasting'.

If being red literally meant the same thing as being sweet-tasting, then a question like the question above would end up sounding like, "That is red, but is it red?" This is NOT an open question. If, instead, the question sounds like an open question, we can take that as evidence that being red is not the same thing as being sweet-tasting, and it is at least conceptually possible that something can be red without being sweet-tasting.

In the realm of value, when a philosopher equates what is good with what is pleasant, or with what is desired, we can ask, "X is pleasant, but is it good?" or "X is desired, but is it good?" If the two terms meant the same thing as some philosophers seem to assert, then this would be like asking, "X is pleasant, but is it pleasant?' or "X is desired, but is it desired?" These are NOT open questions. The fact that these two questions ARE open questions implies that being pleasant (or being desired) is not the same thing as being good.

---

[6] Moore, G.E., *Principia Ethica*, Chapter 4, Section 66, http://fair-use.org/g-e-moore/principia-ethica/chapter-iv, Retrieved 01/03/2017.

[7] Moore, G.E., *Principia Ethica*, Chapter 4, Section 67, http://fair-use.org/g-e-moore/principia-ethica/chapter-iv, Retrieved 01/03/2017.

## Summary

So, the next time somebody comes up to you in a debate and says, "To be good means <insert whatever they think it is to be good>, and suggests that he can prove this by showing you that wherever goodness seems to be found we find <whatever they think it is to be good>, you can accuse them of committing the naturalistic fallacy. The fact that two things are found together does not imply that they are, in fact, the same thing. Furthermore, you can then show that 'good' is distinct from '<whatever they think it is to be good>' by pointing out that "X is good, but is it <whatever they think to be good>' is an open question.

But what can you say if somebody accuses you of committing the naturalistic fallacy?

## Analysis

Let us assume that you are on the receiving end of one of these accusations. Is there anything you can say to defend yourself? Or are you doomed?

If you have actually committed what Moore calls the naturalistic fallacy, you may have a problem. However, we can ask whether the claim that you have committed this fallacy is accurate.

Moore makes a mistake, regarding the definition of terms, that prevents him from seeing a way in which 'good' can be related to a natural property that does not commit the naturalistic fallacy. The person who makes this move is not saying that, because 'good' is found with some other property that they already mean the same thing. Instead, one can assert that the concept of 'good' matches so closely to some other property that we have practical reasons to simply say that they mean the same thing.

## Moore on Definitions

To see how we can generate a response to the accusation of committing the naturalistic fallacy, we start by looking at Moore's ideas about the definitions of terms.

Moore claims that there are two types of terms. Complex concepts identify things that are made up of parts, and define those things by listing their parts and how those parts fit together. The term "horse", Moore argues, can be defined because horses are made up of parts.

However, when we divide these parts into their constituent components, we will eventually get to concepts that are not made of any smaller parts. These simple concepts cannot be defined – specifically because we cannot list their parts and how they fit together.

These simple concepts are "undefinable" on Moore's account. 'Good' is one of those simple, undefinable terms.

> *The most important sense of definition is that in which a definition states what are the parts which invariably compose a certain whole; and in this sense good has no definition because it is simple and has no parts. It is one of those innumerable objects of thought which are themselves incapable of definition, because they are the ultimate terms of reference to which whatever is capable of definition must be defined.*[8]

---

[8] Moore, G.E., *Principia Ethica*, Chapter 1, Section 10, http://fair-use.org/g-e-moore/principia-ethica/chapter-i, Retrieved 01/03/2017.

Moore's theory of definition is quite similar to the ancient Greek concept of matter. Imagine taking a brick of gold and breaking it into parts. You break each part into still smaller parts. There must be a point, according to the ancient Greeks, where you cannot divide the gold any further. There must be a smallest indivisible part – these smallest indivisible parts are what the ancient Greeks called "atoms".

In Moore's theory of definitions, these simple parts cannot be defined. In fact, to say that a word – such as 'good' – is indefinable is the same as saying that the term has no parts.

> *Good, then, if we mean by it that quality which we assert to belong to a thing, when we say that the thing is good, is incapable of any definition, in the most important sense of that word.*[9]

However, there are other ways to define concepts – other relationships that one thing can have to other things other than the relationship of parts to whole.

## More on Definitions

Language is a tool used largely for communication – to cause ideas to appear in the thoughts of the person who receives the communication. A 'definition' is anything that one can do to cause the desired idea to appear in the thoughts of the recipient. One way to define a term is by listing its parts and the relationships between them. However, that cannot be the only way to define a term.

In fact, when it comes to defining the term 'good', the question is: What can we do to get the same idea to show up in the minds of different people when they receive the term 'good' in a communication?

One of the ways which we can define a term – one that refers to a part of a whole – is by listing its relationships to the other parts. Using the concept of a horse – which is a complex concept made up of multiple parts – we can define those parts by describing their relationships to other parts. We can define a molar by describing its positon in the mouth, and each of the other bones by describing their location relative to the other bones.

In fact, some of our terms are strictly relational – they are terms we use to describe the relationships between one thing and another. Consider, for example, the term "sister". You can know everything there is to know about the parts of a horse – its liver, its eyes, its tail, its blood – and none of this will help you to understand the idea that this horse is the sister of some other horse.

One might say that we can define 'sister' in that a sister is a part of a family. However, Moore does not consider the possibility that we can define parts according to what they are a part of. Complex things are defined in terms of their parts, but parts are not to be defined by their relationship to the whole. If terms could be defined according to what they are a part of, then every term would have a definition and there would be no "undefinable terms."

One particular type of relational definition that is particularly useful is a functional definition – where we define something according to the role that it plays in a larger system. One can hand you all of the parts of a mousetrap. However, an understanding of all these parts and even an understanding of how the parts are put together does not allow a person to understand that this contraption has a function – to

---

[9] Moore, G.E., *Principia Ethica*, Chapter 1, Section 10, http://fair-use.org/g-e-moore/principia-ethica/chapter-i, Retrieved 01/03/2017.

trap mice. In fact, any number of devices, having different structures and different parts, can all be properly called a 'mousetrap," having nothing in common but the fact that they are used to catch mice.

There are certainly cases in which it is impossible to define a term. I cannot think of anything I can do to cause my dog (assuming I had a dog) to have an idea of a disjunctive syllogism when I use the term 'disjunctive syllogism' – or even to have an idea of good when I use the term good. I think I can get him to have a concept of the term 'toy'. At least, I can use the term to create behavior that I can predict towards the object I have in mind when I use the phrase 'get your toy' when speaking to my dog.

If we cannot get different people to have the same thought when they receive the term 'good', then we cannot have communication. Somehow, there must be a way to accomplish this – which means that, somehow, there must be a way to define the term 'good'.

One may object to the idea that the definition of a term looks to creating a common idea in the thoughts of those who use it by claiming we have terms that may well lack a common idea. How do we know, for example, that what you see when you see red is the same thing that I see when I see red?

Yet, what matters in the definition of 'red' is not this 'qualia' or the sensation of redness. It is found in the fact that, while we are working together defusing a bomb, and I say, "cut the red wire," you correctly identify the wire that I want you to cut. The qualia of redness is entirely irrelevant to the meaning of the term.

Ironically, this is a place where Moore's naturalistic fallacy actually has an application. A person experiences a certain sensation with all things red and makes the mistaken inference that the term 'red' refers to that sensation. But it does not- because others can be using the same term without having the same sensation. And when one is corrected for having misused the term, it is not on the basis of using it to refer to the wrong sensation. It is because of a failure to identify the correct objects.

The attempt to define 'good' is an attempt to get a common, useful idea to show up in the thoughts of different people when they use the term in a certain set of contexts.

## Ways of Defining Terms

There must be a way to get different brains to have (roughly) the same idea – close enough for communication to take place – when we use a term like 'good'. There must be a way to define the term.

Moore identified three ways of defining a term. I have already discussed the first of these – defining a term by describing the relationships of its parts. The other two are as follows:

> *(1) We may mean merely When I say horse, you are to understand that I am talking about a hoofed quadruped of the genus Equus. This might be called the arbitrary verbal definition: and I do not mean that good is indefinable in that sense. (2) We may mean, as Webster ought to mean: When most English people say horse, they mean a hoofed quadruped of the genus Equus. This may be called the verbal definition proper, and I do not say that good is indefinable in this sense either; for it is*

> *certainly possible to discover how people use a word: otherwise, we could never have known that good may be translated by gut in German and by bon in French.*[10]

These are also sometimes called the stipulated definition and common usage.

With a stipulated definition, an author merely identifies a 'shortcut' – telling readers or listeners that idea is to show up in their thoughts when the author uses a particular term. The author simply asserts, "When I use the term X, I am to be understood to mean . . ."

With the public definition of a term, one makes a claim as to the idea that commonly appears in the thoughts of native language speakers when a term is used in a particular context. Such a claim can be verified or falsified by looking at evidence of the ideas that actually do appear in the thoughts of native speakers of a language when they encounter a term in that context.

Neither of these ways of defining a term are particularly useful in presenting a moral theory.

In the case where a philosopher offers a stipulated definition – where the philosopher simply asserts, "When I use the term 'X', you shall take me to mean 'description of X,' the author cannot say anything useful about the term as it is commonly understood – about the idea that comes into the minds of native speakers on encountering the term. I could stipulate, for example, that by the term 'good' I am to be understood to mean 'produces pleasure'. However, I cannot go on from this to argue, 'Thus, I have proved that what is 'good' is that which produces pleasure." It would be the same as at first stipulating, "When I use the term '5', you are to understand me as meaning 'that which is the sum of 2 + 2", then asserting, "I have thus proved that 2 + 2 = 5". It is true that I have accomplished this result – but I have done so in a way that says nothing meaningful.

In the case of the philosopher who focuses only on the common usage, this is merely a descriptive account of what people think can be said using the term in question – and may differ significantly from what is in fact the case. A lexicographer is charged with noting how a term is used in simply recording these facts – without a regard as to whether the term actually makes sense. If people use the term vaguely or ambiguously, then the lexicographer notes these vague or ambiguous uses.

A sociologist might be interested in describing what people in a culture call good. However, the way that people use the term 'good' is not a measure of what is good as a matter of fact. This is true in the same way that what people call 'flat' is not a measure of what is actually flat. If people are in the habit of thinking that the earth is flat, then the term 'flat' may be defined, in part, in terms of what these people take the shape of the earth to be. However, it does not tell us what the shape of the earth is in fact.

When a philosopher is interested in what really is good, the philosopher is not going to take the way that people use the term as being incapable of error.

> *Do, pray, act so, because the word good is generally used to denote actions of this nature: such, on this view, would be the substance of their teaching. And in so far as they tell us how we ought to act, their teaching is truly ethical, as they mean it to be.*

---

[10] Moore, G.E., *Principia Ethica*, Chapter 1, Section 8, http://fair-use.org/g-e-moore/principia-ethica/chapter-i, Retrieved 01/03/2017.

> *But how perfectly absurd is the reason they would give for it! You are to do this,*
> *because most people use a certain word to denote conduct such as this.[11]*

So, if a person offering a moral theory tries to begin with a definition of "good" – and seeks to offer us either a stipulated definition or a "common usage", the philosopher is providing us with a poor moral argument.

## Reductive Definitions

There is, actually, another type of definition that Moore does not consider. It involves a reduction of one concept to another.

Reduction is a particular blend of stipulation and common usage. Reduction involves taking a common term (e.g., gold, water), adding information that we have acquired about the things to which the term typically refers to in common usage, and stipulating a more precise definition of that term.

For example, we take the common term "water". We combine this with the discovery that much of what is true of water is true in virtue of the fact that the most common substance in that to which the term 'water' commonly referred is made up of a combination of one oxygen atom with two hydrogen atoms – or $H_2O$. We then stipulate that we are going to adjust our language so that the term 'water', henceforth, will refer to '$H_2O$'.

Moore's "naturalistic fallacy" objection applied to this type of reduction would go like this:

*Look, what you are doing is that you are looking at water and seeing that, wherever you find water, you also find $H_2O$. From here, you are making a rather remarkable mistake – remarkable in the sense that it is quite obvious and a wonder that you do not recognize it – of inferring that, as a result, "water" must mean the same thing as $H_2O$. Yet, clearly the two terms do not mean the same thing. We can show this by the fact that, 'This glass contains water, but does it contain $H_2O$?' is an open question – a person can wonder, without confusion, whether the answer is 'yes' or 'no'. Thus, I refute your attempts to define 'water' as '$H_2O$'.*

However, we are not, in fact, making the mistake of inferring from the fact that we find $H_2O$ wherever we find water that the two terms mean and have always meant the same thing. In fact, we can confidently assert that they have not always had the same meaning – that, until recently, the term '$H_2O$' had no meaning whatsoever. However, now that we have an understanding of $H_2O$, we can propose defining 'water' as '$H_2O$' – we can stipulate a new definition – which preserves the truth value of the vast majority of claims made under the old definition of 'water', but also allows us to speak more precisely and accurately, explaining why those truths are true and adding a lot of new information to this old term.

The same can be said for reducing 'good' – as I would argue it should be reduced – to 'that which fulfills the desires in question'. When we make this reduction, we can explain much of what is known about 'good', throw out a bunch of nonsense, and talk about the goodness of things much more efficiently than we could under the old, vague definition.

---

[11] Moore, G.E., *Principia Ethica*, Chapter 1, Section 11, http://fair-use.org/g-e-moore/principia-ethica/chapter-i, Retrieved 12/25/2016.

A reduction is not a common type of stipulation like that which we objected to in the previous section. One of the characteristics of a reduction is that we are redefining a term in a way that not only captures many of the things that are known about the term in common language – but explains much about what is known. We can reject a reduction if it is too extreme – if it captures and explains too little of what people commonly claim to be true of the original term.

Of course, when people used the term before these discoveries were known, they clearly could not have meant by the term that which it is now being reduced to. For thousands of years people spoke of water without having any idea of H2O. The question, "This cup contains water, but does it contain H2O?" remains an open question. At least, it remains an open question during the period of transition before the reduction takes hold in the public mind and becomes "second nature" to the speakers of that language.

This fallacy is also heavily discussed in the philosophy of mind, where some people use it to object to the claim that mental states cannot be reduced to physical states of the brain. The person using this objection states, "I can understand fear as a mental state, but I do not understand fear as a state of the brain. In fact, people have thought about fear for centuries without thinking of it as a brain state. Thus, a mental state cannot be a brain state." However, the fact that people for centuries have talked about mental states without thinking of them as brain states no more proves that they are not brain states than the fact that people have talked about water without talking about H2O or carbon without talking about atoms having 6 protons in their nucleus does not prevent us from claiming that this is what water and carbon is.

## Conclusion

In summary, when you wish to use the accusation of "naturalistic fallacy" against another person, you are looking for it to be the case that the person you are using it against has asserted that, because two properties occur together, they are inferring that the terms referring to those things are referring to one and the same thing. They are asserting, from the fact that 'redness' and 'sweetness' occur together, that 'redness' is the same as 'sweetness'. Or that because 'goodness' is commonly found with 'pleasure' that 'goodness' is the same as 'pleasure' (or some other natural property). This is an invalid inference. That the two concepts are, in fact, distinct can be shown using the Open Question argument.

If you are charged with committing the naturalistic fallacy, you should look at whether the accusation is accurate. It may well be the case that what you are doing is not committing the naturalistic fallacy, but attempting to reduce one term ('good') to another ('that which is such as to fulfill the desires in question') in the same way that scientists reduced the term 'water' to 'H2O'. You are not saying that people have for all time recognized that these were the same thing and used the terms accordingly. You are proposing that if we substitute one term for another we can explain many of the ways in which the original term was used – explain why many of the true propositions using the original term are true – and begin using the original term much more precisely and much more accurately.