Essential Kind Membership and Possible Worlds or Could Aristotle Have Been a Paper Clip?
by Eileen Walker

1. Introduction: conflicting intuitions in philosophy[1]

No one *really* thinks that Aristotle could turn into a paper clip or the number 17, or that he could have been any of these things. Our intuitions tell most of us that Aristotle is – was – a human being, and human beings don't undergo these kinds of change. The problem is to explain philosophically *why* this is so. In the case of factual intuitions, it's easy. We go out and do some research. We used to have an intuition that the sun moved round the earth, but we found out we were wrong. But *philosophical* intuitions are tricky and can divide opinion quite sharply. There are notoriously divergent metaphysical intuitions about such subjects as free will, the existence of God and causal necessity. When this happens, side A, which supports the intuition, tries to explain it and provides arguments in its favour to show why it is justified. Side B tries to explain the *appeal* of the intuition but provides arguments to show that it is fundamentally mistaken – they try to 'explain it away'.

2. Essential Kind Membership
The intuition divide here concerns the identity and persistence conditions of an individual through changes in its properties. An 'individual' is anything that can be counted: an object, an entity a 'particular', to use Aristotelian terminology. What is at issue is whether it is possible for an individual *x,* that starts out as one kind of thing, to metamorphose into a radically different kind of thing - to become a member of a totally distinct kind. Can Aristotle, who comes into existence as a *human being*, change into a giraffe, a parsnip or the number 17? Could he have *been* one of these things in another life?

Side A in the debate says 'no'. They support a claim that I call 'Essential Kind Membership' - EKM:

Def: Essential Kind Membership
  i. Nothing can change from being an F to being a not-F and still continue to exist.
  ii. Nothing that is *actually* an F can have spent its entire existence as a not-F.

On the A side In favour of EKM are Aristotle himself and contemporary philosophers, Kit Fine, EJ Lowe, Saul Kripke and David Wiggins. Call them the neo-Aristotelians. [Oddly, some contemporary essentialists, Alexander Bird and Brian Ellis, reject the notion of EKM.]

Side B rejects EKM because even though most people take for granted that a human being is essentially a human being, they see no good arguments in favour of it, and have different intuitions as to how the world works.

On the B-side are David Hume, and in our own day David Lewis and Penelope Mackie. Call them the 'neo-Humeans'. However, these are neo-Humeans who engage in metaphysics [unlike the logical positivists, and empiricists such as Karl Popper who dismiss all metaphysics as rubbish]. So how do 'Possible Worlds' come into this?

3. What are 'Possible Worlds'?
a. Clarifications
First, 'possible worlds' are not the physicists' 'multiverses' e.g. in cosmology or quantum physics. 'Possible

---

1   A health warning: this piece was written as a talk, to provoke discussion. There are some inevitable over-simplifications that would be inappropriate in a more fully developed paper.

worlds' in philosophy are a logical tool. They are useful because they reflect our intuitions regarding modality, that is, regarding what is necessary, possible or impossible. Classical logic until early in the 20C dealt with what *is* the case and what is *not* the case. It was not designed to deal with how things might be or might have been. New modal logics were invented, but there are ideas that cannot be expressed in them, and the notation can be very complicated. So 'possible worlds' provided a means of assigning truth values to conditional and counterfactual statements.

b. Counterfactuals (how things might have been)

Intuitions about how things might have been may be expressed using 'possible worlds' talk. The statement: 'if I'd left home 5 minutes earlier, I'd have caught the 4.57 train' can be expressed by saying that there is a possible world where I left home 5 minutes earlier and did catch the train. The statement: 'pigs might fly' is expressed by saying that there's a possible world where pigs do fly. You may of course think that there is no possible world where pigs fly. This is where the modal metaphysics comes in.

c. Modal definitions

- Necessity, possibility and impossibility are *interdefinable* – think about it. Of course, anything which is *actual* must be *possible*.
- A proposition is *necessary* if it is true in *all* possible worlds. *Example*: '2+2 = 4'.
- A proposition is *impossible* if it is true in *no* possible world. *Example*: 'This shirt is both blue and not blue'.
- A proposition is *possible* if it is true in at least one possible world. *Example*: 'My car is grey'.
- 'Essential' and 'necessary' for the purposes of Possible Worlds come down to the same thing. This view is criticised by Kit Fine (see note 2, section 7a).

We now look at how Side A and Side B use Possible Worlds to address the Essential Kind Membership question. I start with Side B, the opponents of EKM.

4. Against EKM: The neo-Humean or David Lewis view of Possible Worlds (*On the Plurality of Worlds*, 1986)

This is the 'Side B' view which 'explains away' our intuition that if Aristotle is a human being, he could not turn into a parsnip or a paperclip. For David Lewis, the most eminent, gifted and startling philosopher in the neo-Humean tradition, all possible worlds are *real*, but only one world is *actualised* – that is the one we are in. Our job is to work out which world that is by excluding all the *unactualised* worlds. Impossibility and possibility are cashed out in terms of *remoteness* and *proximity*. Remote worlds would be those with spaceships exceeding the speed of light, giraffes melting into puddles and philosophers turning into paperclips. Nearby worlds are those where, for example, this talk took place in the Lecture Theatre, or where I fell sick and Peter gave a talk instead.

This is a perfectly coherent view if one subscribes to the Humean idea that anything which is conceivable is possible, so long as it is not ruled out by the laws of logic. There is no logical contradiction in the idea of a melting giraffe or a philosophising paper clip. Add to that the view that individual objects like Aristotle are nothing more than 'bundles' of properties, none of which is necessary to the bundle as a whole. If one property in the bundle changes, the bundle is no longer the *same* bundle. So in each possible world where he has a property change, the bundle which is Aristotle becomes a *different* bundle – a 'counterpart' of the actual world Aristotle. The Aristotle who is a fisherman in a nearby possible world is a 'counterpart' of Aristotle the philosopher, as is the paperclip in the remote world. Any property-bundle combination is possible unless it is blatantly contradictory.

The advantage of this view is that it enables a neo-Humean empiricist to accept that we attribute properties and causal powers to natural kinds in the world as we observe them, but argue that these are not necessary. Take a simple example: a chemical substance like *water* has the causal power to dissolve salt. In a different

possible world, it might not have that power. Salt might remain suspended in water, like sand does. Or *water* might have a different chemical constitution.

5. In favour of EKM: The neo-Aristotelian view

There are three distinct responses to the view that anything might have been anything. I'll deal with two of them now. (The third is that of Kit Fine - see section 7a, fn. 2.)

a. Saul Kripke, b.1941 (*Naming and Necessity* 1980)

Kripke is one of the key developers of modal logic and the notion of a 'possible world', but his conception is quite distinct from that of David Lewis. Possible worlds are states our *actual* world – our universe – might be in or might have been in. We take our *actual* world Aristotle, *that very man,* and then evaluate what might have happened to *him*, had things gone differently in our *actual* world. We can then say that *he* might not have been taught by Plato, *he* might not have been the teacher of Alexander the Great, and even that *our Aristotle* might not have gone into philosophy at all and been a fisherman instead (for obvious reasons, we can't say that he might have been an astronaut). We may use the language of 'possible worlds' to express such thoughts, though we need not. What we do *not* do, according to Kripke, is cast a flashlight round possible worlds to see which individual has the bundle of properties that most closely resembles *our* Aristotle. That is entirely the wrong picture. On the Kripkean view, the Aristotle who became a fisherman and did not teach Alexander is the *very same* Aristotle with slightly different properties, not a 'counterpart' of him. A counterpart of Aristotle would not *be* Aristotle, but a distinct individual.

If we heed this message, counterfactual speculation about Aristotle has to be anchored in the *man himself.* This is where David Wiggins comes in.

b. David Wiggins, b. 1933 (*Sameness & Substance Renewed* 2001)

Wiggins does not argue in terms of 'possible worlds' though he accepts the Kripkean view. Wiggins argues that before we can evaluate what changes are possible and might have been possible for an individual, we first need to have some conception of what *kind* of thing that individual is. The *kind* to which an individual belongs determines its identity – it determines *what* it is. And what it *is* determines its *persistence* conditions – the changes it can undergo while still remaining the *same object*. The persistence conditions of an individual are *kind-specific* and stem from empirically-discovered essential properties. Though Wiggins does not give examples, essential properties would include having a particular chemical constitution, a particular atomic number, a particular biological origin and genomic profile etc, etc.

Recall that Side A is trying to defend Essential Kind Membership:
  i. Nothing can change from being an F to being a not-F and still continue to exist.
  ii. Nothing that is *actually* an F can have spent its entire existence as a not-F.

Wiggins has the Aristotelian notion of a *substantial kind.* It lends support to the *first* plank of EKM.

> *Def:* a kind *K* is a 'Substantial Kind' *if and only if* it is such that an object *x* that belongs to *K* at *any* time in its existence belongs to *K* at *all* times in its existence.

Substantial Kind membership can be summed up in the 'continuity' slogan 'once an F always an F' (where 'F' indicates membership of an appropriate Substantial Kind). A consequence of Substantial Kind membership is that an individual *x* comes into existence *as* a member of a particular kind, with the set of persistence conditions distinctive of that kind, and goes out of existence as a member of that *same* kind. Wiggins' illustration is the Biblical legend of Lot's wife turning into a pillar of salt. Once she is no longer a *human being*, *she* ceases to exist and a totally distinct entity replaces her. Applying the Continuity Principle helps us distinguish between Substantial Kinds like *human being* and what Aristotle and Wiggins call

'accidents', which are contingent and often temporary states that members of Substantial Kinds can be in. We can say 'once a human, always a human', or 'once a parsnip always a parsnip'. We can also say 'once a piece of gold always a piece of gold'. But we can't say 'once a brain-surgeon always a brain-surgeon'. A brain-surgeon does not cease to exist if he decides to re-train as an accountant.

A word of caution: a caterpillar does not cease to exist when it pupates and turns into a butterfly, but being a caterpillar is not an *accident*. Phase state changes are an essential part of the organism's life-cycle, should it live so long – they are part of the persistence conditions of members of the Substantial Kind *Lepidoptera*. Phase states are part of the natural persistence conditions of chemical substances too, so there is no change of substance when a block of ice melts into a puddle of water.

If there are Substantial Kinds, as Wiggins suggests, we have *some* grounding for the claim that Aristotle cannot turn into a paperclip. But we are not home and dry yet. Penelope Mackie has what is regarded as a strong argument against Essential Kind Membership. Back to Side B, the opposition.

6. Penelope Mackie against EKM

Recall that Essential Kind Membership makes two claims.

  i. Nothing can change from being an F to being a not-F and still continue to exist.
  ii. Nothing that is *actually* an F can have spent its entire existence as a not-F.

If natural kinds are regarded as Substantial Kinds then we have some grounds for accepting the first plank of EKM. But now the 'killer blow'.

Mackie points out that there is a logical gap between the two planks of the EKM claim. The first does not entail the second. One can accept the first while rejecting the second. Although Aristotle cannot *change* from a human being into a centipede, a puddle of water or the number 17, it does not *logically* follow that he could not have spent his entire existence as one of these other things instead.

Now recall again the Substantial Kind claim which sustains the first plank of EKM:

> *Definition:* a kind *K* Is a Substantial Kind *iff* it is such that if an object *x* belongs to *K* at *any* time in its existence then *x* belongs to *K* at all times in its existence.

What we need to support plank ii) of EKM is an *Essential Kind* claim, which substitutes *worlds* for *times*:

> *Definition:* a kind *K* is an Essential Kind *iff* it is such that if an object *x* belongs to *K* in *any* possible world – which will of course include the *actual* world – then *x* belongs to *K* in *all* possible worlds where *x* exists.

If Substantial Kinds such as *human being* are Essential Kinds, then anything that is a human being in the *actual* world will be a human being in all possible worlds where it exists. That's what most of think anyway – it's the intuition we started out with. But how to argue for it?

One way is suggested by Mackie herself:

> 'if there is *any* kind of thing such that, necessarily, everything that belongs to that kind belongs to it essentially, then if Aristotle does *not* in fact belong to that kind, he could not have done so.' (2006: 165) For example, if it is the case that every number is essentially a number, then if Aristotle is a number in any world, he must be a number in all possible worlds, including the actual world. Since he is clearly not a number in the actual world, he could not have been a number in any world.

Even if we cannot say what Essential Kinds Aristotle *does* belong to, there might be Essential Kinds from which he is necessarily *excluded*. And Mackie in fact suggests that there might *be* such general Essential Kinds - *material object, abstract object, event, number* - that we can rationally accept, even if we reject essentialism. So if Aristotle is not a football match, a mathematical theorem or the number 17 in the *actual* world, then he could not have been any of these things. But… if *material object is* an Essential Kind, we cannot exclude Aristotle from it because he too is a material object – in which case he could still have been a paperclip! It looks as if the B-side is winning. [The Members' Day talk stopped at this point.]

7. Defending EKM against Mackie
a. Kit Fine: essence, identity and necessity

We may accept Mackie's point that there is a *logical* distinction between a Substantial Kind and an Essential Kind, but we should regard it as *irrelevant*. Since we are dealing with the identity of individuals in relation to the kinds to which they belong, what is at issue here is not *logical* necessity/possibility but metaphysical necessity. Kit Fine (2005: 7) distinguishes between these two varieties of necessity/possibility as follows:

> A logical necessity has its source in logical form; it is true, or necessary, in virtue of its logical form… A metaphysical necessity has its source in the identity of objects; it is true, or necessary, in virtue of the objects with which it implicitly deals.

> If this account is to have any explanatory value, we must distinguish between the identity or 'essence' of an object and the properties it has as a matter of necessity.

If Fine is right, logical possibility is not the relevant form of modality with which to evaluate counterfactual possibilities for a particular x.[2] We need instead to attend to the metaphysical possibilities which flow from the identity and nature of x itself. What is metaphysically possible for x can only be evaluated once x actually comes into existence. If x comes into existence in the actual world as a member of the Substantial Kind *Homo sapiens,* call him '*A*', then the identity of *A* is fixed *as* a member of *Homo sapiens* for any other world where *A* exists. On *A*'s essence as a human being depends what is actually and counterfactually possible for *A*. The claim that *one and the same* individual, *A*, could be a member of *Homo sapiens* in one world and a parcel of gold or a paperclip in another world is a claim that $A \neq A$, a blatant violation of the necessity of identity.[3] If x is a member of *Homo sapiens* and y is a parcel of gold, then necessarily x ≠ y, whether in this world or in other worlds.[4]

This picture does give us Essential Kind Membership. It also has the consequence that any Substantial Kind is an Essential Kind. The steps are as follows:

---

2   The last sentence of the quotation is a reference to Fine's 1994 paper, where he notes the asymmetry between the essential and necessary properties of an object, which a 'possible worlds' analysis of necessity fails to bring out. For example, it is *essential* to the singleton set of Socrates that it has Socrates as its sole member. But it is not essential to Socrates that he is the only member of the singleton set of Socrates, although Socrates exists at every world where his singleton set exists. Socrates is necessarily part of singleton Socrates because of the essential nature of singleton sets (not because of the nature of Socrates). So facts about essences are the *grounds* not the consequences of facts about metaphysical necessity, i.e. possible worlds facts.

3   The only way to avoid this logical contradiction would be to opt for some version of Lewis's counterpart theory, which Mackie rejects, despite sympathies with Lewis.

4   Wiggins' says: 'Facts about the identity and difference of individuals are part of the necessary structure of reality and are completely invariant across possible worlds. If the objects *a* and *b* are anywhere distinct, then there is no possible set-up where they are the same. If *a* and *b* are the same anywhere, there is no possible set-up where they are distinct.' (*S&SR*: 117)

1. A Substantial Kind *K* is such that if any particular *x* belongs to *K* at any time in its existence, then *x* belongs to *K* at all times in its existence.
2. Since *x* comes into existence *as* a member of *K,* its existence and persistence conditions for *all* possible worlds are determined by its essential properties as a member of *K*.
3. It follows that if *x* is a member of *K* in *any* world, then *x* is a member of *K* in *all* possible worlds where *x* exists.
4. So if *K* is a Substantial Kind, then *K* is an Essential Kind.

b. Back to Mackie's argument

Now recall Mackie's argument: 'if there is *any* kind of thing such that, necessarily, everything that belongs to that kind belongs to it essentially, then if Aristotle does *not* in fact belong to that kind, he could not have done so' (2006: 165). Mackie suggests that there might be such Essential Kinds: *material object, abstract object, event, number*. Then there would be categories from which Aristotle was necessarily excluded. If he is not a snowstorm, a mathematical theorem or the number 17 in the actual world, then he could not have been any of these things. Mackie's reason for accepting these very general Essential Kinds seems to be that it makes her own position less counterintuitive and implausible.[5]

However, it simply makes Mackie's position inconsistent. Why does she choose *these* particular categories and not others? Clearly, these very general Essential Kinds are subject to what we may call an 'Exclusion Principle', and this can be argued: for example, nothing can be both a material object and an abstract object. But Mackie does not argue this, perhaps because it opens the door to other essentialist exclusions, such as those we see in the natural world.[6] Applying the Exclusion Principle, we may say that nothing can be both a living thing and a non-living thing. If Aristotle is a human being he is a living thing. He cannot therefore be a non-living thing such as a paperclip in this or any other world. But we can also apply Mackie's own argument against her. If *non-living thing* is an Essential Kind, then if Aristotle were a non-living thing in any world, he would be a non-living thing in the actual world, which clearly he is not.

Mackie's argument hinges on what we are rationally obliged to accept. She does not explain why it is more rational for us to accept rather than reject the view that a human being, *contra* our understanding of how the world operates, might have been a centipede, a parsnip or a paperclip. Given EKM, the counterfactual possibilities for Aristotle are grounded in his identity and nature in the actual world, where he is a human being, a member of the Essential Kind *Homo sapiens*. This constrains what is metaphysically possible for Aristotle in other worlds. Metamorphosing into a paperclip is not an option (a conclusion that Mackie would accept). But, as we have seen, nor is the possibility of his having been a paperclip for his entire existence.

Bibliography
Fine K. 1994, 'Essence & Modality' *Philosophical Perspectives* 8 (ed. J. Tomberlin): 1–16.
_____2002, 'The Varieties of Necessity', reprinted in Fine 2005.
_____2005, *Modality and Tense: Philosophical Papers,* OUP.
Kripke S. 1980, *Naming and Necessity*, enlarged edition, Blackwell. (*N&N*)
Lewis D. 1986, *On the Plurality of Worlds,* Blackwell.
Mackie P. 2004/2006, *How Things Might Have Been*, OUP.
Wiggins D. 2001, *Sameness and Substance Renewed*, Cambridge University Press.

---

5   Mackie claims to be agnostic as to whether there *are* such essential kinds, but thinks that to deny such a possibility might seem dogmatic (2006: 166).

6   For example, no chemical element can have more than one atomic number.