

# A 'TRINITARIAN' THEORY OF THE SELF

HOWARD ROBINSON

*Central European University, Budapest*

**Abstract.** I argue that the self is simple metaphysically, whilst being complex psychologically and that the persona that links these moments might be dubbed 'creativity' or 'imagination'. This theory is trinitarian because it ascribes to the self these three 'features' or 'moments' and they bear at least some analogy with the Persons of the Trinity, as understood within the neo-platonic, Augustinian tradition.

## I. INTRODUCTORY REMARKS

The theory of the self that I want to defend in this paper is trinitarian because it ascribes to the self three 'features' or 'moments' - with these expressions used in the most neutral sense possible - and these three bear at least some analogy with the Persons of the Trinity. In fact, the theory might be thought of as having more direct affinity with neo-Platonism than with the Christian doctrine, though that there are important parallels there, especially in the Augustinian tradition, is very well known.

In order to construct this picture of the self, I shall, in some sections of this paper, draw on arguments that I have deployed elsewhere, and I apologize for a certain amount of repetition of previously published themes. I shall do this only to the extent necessary to make the argument presented here clear and plausible in its own right.

## II. OUTLINE OF THE POSITION

The three features of the self are as follows. First, the self is a simple entity. This is a view that I have defended in various places, most recently in Robinson (2011). Second, there is the obvious complexity of the Self,

as revealed in the fact that we all have many mental states and processes. This raises the obvious problem of how a simple thing can also be complex. The third feature, which is connected with this latter problem, is the process or manner by which the essentially simple self expresses itself in the complexity of its psychology. I shall argue that the human subject is equally present in all these features or modes of its existence, that is, as a simple metaphysical entity, as the complex psychology in which it is expressed, and in the process of expression that leads from the simple subject to the complicated psychology.

### III. THE SIMPLE SELF

The argument for the simplicity of the self is currently under-discussed, although it seems to me to be an argument of considerable importance. I shall try to rehearse the argument with as much brevity as the need for persuasive detail will allow.

There is a long tradition of arguing that the identity of persons over time is not a matter of convention or degree in the way that is the case for complex physical objects – such as Theseus's ship. There is something absolute – all or nothing – about one's being numerically the same person at 70 as at 7.

But this intuition is controversial and does not command universal, or even general, assent. Growth, aging, and especially radical changes in personality due to accidents or diseases are claimed to make one into 'a different person'.

I think that the issue can be made sharper and clearer, however, if one moves from considering identity through time to the rather less familiar matter of identity under counterfactual circumstances, especially those concerning origin. Instead of asking whether Theseus's ship was the same object when half its planks had been replaced, we ask whether it would have been the same ship if it had been constructed with different materials in the first place. So we are not considering changes within its life as a boat, but possible differences at its origin. Thus we are considering *counterfactuals of origin*, that is, things that might have been different at the beginning of the existence of an object. (Such things are *counterfactuals* because they state how things *might* have been, not how they, in fact, were.) We would probably agree that if the ship had been made not of wood but of gold, it would not have been the same ship at all. But if it had been made of, say, 10% different planks and 90%

the same ...? This thought experiment can be duplicated for any complex physical object. One might be tempted to follow Williamson's epistemic line and say that there must be a fact, just one that is unbeknownst to us (Williamson 1994). I – and many others – have argued against that as a treatment of vagueness elsewhere (Robinson 2008-9). I am going to assume that the correct response to such things is to say that there is no fact of the matter about whether it would or would not have been the same in the borderline cases. Once the story has been told about such and such differences, those are all the real facts. There is more or less overlap of constitution, but what, if anything, one says about identity is a matter of choice. As I hope to show, a similar treatment cannot be meted out in the case of persons, when it comes to these counterfactual cases, even though it looked as if it could in the case of identity through time.

Let us try to apply the same thought experiment to a human being. Suppose that a given human individual – call him Jones – had had origins different from those which he in fact had such that whether that difference affected who he was is not intuitively obvious. We can approach this by imagining cases where it seems indefinite whether what was produced was the same body as Jones in fact possesses. What would count as such a case might be a matter of controversy, but there must be one. Perhaps it is unclear whether Jones's mother would have given birth to the same human body if the same egg from which the Jones body came, had been fertilized by a different though genetically identical sperm from the same father. Some philosophers might regard it as obvious that sameness of sperm is essential to the identity of a human body. In that case, imagine that the sperm that fertilized the egg had differed in a few molecules from the way it actually was; would that be the same sperm? If one pursues the matter far enough there will be indeterminacy which will infect that of the resulting body. There must therefore be some difference such that neither natural language nor intuition tells us whether the difference alters the identity of the human body; a point, that is, where the question of whether we have the same body is not a matter of fact.

These are cases of substantial overlap of constitution in which that fact is the only bedrock fact in the case: there is no further fact about whether they are 'really' the same object.

My claim is that no similar overlap of constitution can be applied to the counterfactual identity of minds.

To see why this is so, imagine the case where we are not sure whether it would have been Jones' body – and, hence, Jones – that would have

been created by the slightly modified sperm and the same egg. Can we say, as we would for an object with no consciousness, that the story 'something the same, something different' is the whole story: that overlap of constitution is all there is to it? For the Jones body as such, this approach would do as well as for any other physical object. But suppose Jones, in reflective mood, asks himself: 'If that had happened, would I have existed?' There are at least three answers he might give to himself. (i) 'I either would or would not, but I cannot tell.' (ii) 'In some ways, or to some degree, I would have, and in some ways, or to some degree, I would not. The creature who would have existed would have had a kind of overlap of psychic constitution and personal identity with me, rather in the way there would be overlap in the case of any other physical object.' (iii) 'There is no fact of the matter whether I would or would not have existed: it is just a miss-posed question. There is not even a factual answer in terms of overlap of constitution.'

The second answer parallels the response we would give in the case of bodies. But as an account of the subjective situation, I claim that it makes no sense. Call the creature that would have emerged from the slightly modified sperm, 'Jones\*'. Is the overlap suggestion that, just as, say 85% of Jones\*'s body would have been identical with Jones' original body, and about 85% of his psychic life would have been Jones'? That it would have been like Jones' - indeed that Jones\* might have had a psychic life 100% like Jones' - makes perfect sense, but that he might have been to that degree, the same psyche - that Jones '85% existed' - makes no sense. Take the case in which Jones and Jones\* have exactly similar lives throughout: which 85% of the 100% similar mental events do they share? Nor does it make sense to suggest that Jones might have participated in the whole of Jones\*'s psychic life, but in a rather ghostly only-85%-there manner. Clearly, the notion of overlap of numerically identical psychic parts cannot be applied in the way that overlap of actual bodily part constitution quite unproblematically can.

It is important to notice how the identity across counterfactuals of origin case differs from that of identity through changes across time. It concerns what one might call *empathetic distance*, which is essential to the problematic nature of identity through time but irrelevant in the counterfactual case.

Suppose that my parents had emigrated to China whilst my mother was pregnant with me, and that, shortly after my birth, both my parents had died. I was then taken in by Chinese foster parents, lived through

the revolution and ended up being brought up in whatever way an alien would have been brought up in Mao's China. None of this person's post-uterine experiences would have been like mine. It seems, on the one hand, that this person would obviously have been me, and, on the other, that it is utterly unclear what kind of empathetic connection I can feel to this other 'me'. If I ask, like Jones, 'would this have been me?', I am divided between the conviction that, as the story is told, it obviously would, and a complete inability to feel myself into the position I would then have occupied. This kind of failure of empathy plays an important role in many stories that are meant to throw doubt on the absoluteness of personal identity. It is important to the attempt to throw doubt on whether I am the same person as I would become in fifty years time, or whether brain damage would render me 'a different person' in more than a metaphorical sense. It is also obviously something that can be a matter of degree: some differences are more empathetically imaginable than others. In all these cases our intuitions are indecisive about the effect on identity. It is an important fact that problems of empathy play no role in the counterfactual argument. The person who would have existed if the sperm had been slightly different, could have had as exactly similar a psychic life to mine in as exactly similar environment as you care to imagine. This shows the difference between the cases I have discussed and the problematic cases that involve identity through time. In those cases the idea of 'similar but not quite the same' gets empirical purchase. My future self feels, in his memory, much, but not all, of what I now feel. In these cases, overlap of conscious constitution is clearly intelligible. But in the counterfactual cases, imaginative or empathetic distance plays no essential role, and the accompanying relativity of identification gets no grip.

We have considered (ii), the option that its identity might be a matter of degree and rejected this. But what about (iii), the suggestion that there is no fact of the matter whether I would or would not have existed? This boils down to the thought that there is no firm difference between qualitative similarity and numerical identity: degree of similarity of personal history is the only fundamental relation between the two cases. After all, we have a strong feeling that there must be more in the case of bodies, yet we seem to have been forced to accept that this is not the case. Could our sense that there must be such a distinction in our own case be an illusion? Is that conception of the self which makes us feel so sure that someone physically just like me but with a somewhat different origin

either is me or is not, something that needs ‘deconstructing’, after the fashion of Derrida, Nietzsche or Hume?

I do not think that the idea ‘just like me but the idea of whether it would be me or not has no content’ can be made acceptable. Whereas in the case of physical objects we can see, after a little thought, that the qualitatively similar gives us all we thought we needed by talking about particulars, it will not do this in the case of minds.

Consider the following example. Suppose you discover that, in the very early stages in the womb, you were one of twins, but that the other did not develop, and that it could have easily happened the other way round; the other would have survived and you died in the first few days. The similarity between you as survivor and your twin, had he survived in your stead, both in genetic endowment and environmental circumstances and subsequent experience, could have been almost complete. Nevertheless, there is no sense that, on reflection, it makes no serious factual difference, concerning your own fate, which of the two survived. Just as it is true that, if your parents had never met, then you would not have existed, equally, if the other bundle of cells had developed instead of yours, you would not have existed. This is, in no sense, a matter of decision, convention or degree.

So it would seem that we can conclude that the only possible answer to the question which I supposed Jones to have asked himself above, ‘if that had happened, would I have existed?’ is (i), ‘I either would or would not, but I cannot tell. If there is a real fact, independent of our convention or decision, in this case, then it shows that counterfactual identity facts are real facts in the case of minds, in a way that they may not be for physical objects.’<sup>1</sup>

One might respond to this argument by claiming that the difference between the twins is quite clear: they come from different physical sources and so have different bodies. This is true but misses the point of the story. If I am convinced that I would not have existed, this does not

---

<sup>1</sup> I have not discussed or allowed for David Lewis’s notorious modal realism. According to Lewis, every possibility represents a completely different spatio-temporal system. So the sentence ‘I might have had a fried egg for breakfast yesterday’ (though I did not) is made true by the existence of a universe spatio-temporally unrelated to this one in which someone otherwise just like me (my ‘counterpart’) did have a fried egg for breakfast on the parallel day. On this view, in the most basic sense, *nothing at all* sustains counterfactuals, because all other possibilities are realized in counterpart entities, which, strictly speaking, are different things from the objects in the world we inhabit.

mean simply that another body would have existed. It states a further fact, even if it is one that depends or supervenes on the identity of the physical body. But then we have to see how this further fact copes with the indeterminate cases. If there is a real fact in the twin's case, and we cannot answer the indefinite cases in way (ii), with some 'more or less' answer, how is one to move from robust fact to there being no fact at all? Once it is established that there really is something at stake, the nihilist option does not seem to be available. If there is a real issue for clear cases, there will be an indeterminacy as to when unclear cases start. There will be an analogue of higher order vagueness here: there is no clear move from clear fact to no fact.

#### IV. HOW DOES THOUGHT DERIVE IN OR FROM THE THINKING SUBJECT?

The argument of the previous sections attributes to the conscious subject a unity and simplicity, but one might wonder how something as complex as a human subject can be a simple entity. People have a variety of faculties and capacities, and an almost unlimited number of memories, beliefs, desires, etc.; what does it mean to say that such an entity lacks parts or composition?

The modern theory of thought and active expression in general is computational. On this picture, the complexity of our mental life derives from a complex source, namely the programme that governs the computation and the innate concepts of a "Language of Thought". But the computational theory of thought has many problems. They all stem, I think, from the fact that the computational theory of mind is necessarily a syntactic theory of mind, and this is, in effect, a form of epiphenomenalism. According to STM, the causal efficacy of thinking is due entirely to the physical structure of the Language of Thought, not in any direct way to the meaning or propositional content of the symbols that constitute the language. Searle in his famous Chinese Room argument showed that this is not sufficient to constitute conscious understanding (Searle 1980). There is also a large literature now which convincingly defends the view that there is a phenomenology of thought and understanding. (See, for example, Bayne and Montague 2011.) Fodor does not believe that understanding has anything to do with consciousness, but this is not plausible. Even if a non-conscious being could think as we think, it is not plausible that our consciousness

is epiphenomenal in relation to the results of our thinking. The idea that, in a thoughtful discussion on a complicated philosophical topic, one's consciousness has nothing to do with how one replies to one's interlocutor is no more plausible than the theory that how pain feels has nothing to do with how one reacts, or that the conscious visual experience of a tiger approaching has nothing to do with one's running away. But functionalism is the only physicalist game in town for consciousness, as it is in the form of the computational theory, for thought. It plainly will not do in either case: the situations are quite parallel. Searle once stood almost alone amongst the leading figures in the philosophy of mind in affirming this forcefully, but he has now been joined by Thomas Nagel. Nagel agrees that the syntactic-physicalist approach to thought, that tries to abstract it from both consciousness and the real efficacy of meaning and understanding, is hopeless.

I shall assume that the attribution of knowledge to a computer is a metaphor, and that the higher-level cognitive capacities can be possessed only by a being that also has consciousness (setting aside the question whether their exercise can sometimes be unconscious). That already implies that those capacities cannot be understood through physical science alone, and that their existence cannot be explained by a version of evolutionary theory that is physically reductive. (2012: 71)

The metaphor of the mind as a computer built out of a huge number of transistor-like homunculi will not serve the purpose, because it omits the *understanding* [italics added] of the content and the grounds of thought and action essential to reason. (2012: 87)

This has the consequence that, contra both Fodor and Dennett, the human mind is a semantic engine as well as a syntactic one. How are we to understand the idea that meaning as well as grammatical form drives our intellectual life? Fodor says that the STM is the only game in town, and, from a physicalistic point of view, that is true. But he also says of the computational model 'the mind does not work that way' (2001): it works only for those processes that are specifically modular, not for the 'common workspace' in which the modules are brought together. This latter claim of Fodor's is controversial, but it is consonant with reservations that Chomsky originally had about the use of his theory of language. As Mark Baker (2011) points out, this lacuna in the scientific explanation of thought was indicated by Chomsky fifty years ago. Chomsky divided language into three elements; the lexicon, syntax and the Creative Aspect



of Language Use. His theory, he claimed, had nothing to say about the last. This is, in part, at least, because understanding is not driven by syntax alone. This would suggest that the Creative Aspect of Language Use, and, hence, the development of thought, when that involves more than formal inferences, but also depends on our grasp of meanings and our understanding of our own projects, depends on something more than features of the neural/computational machinery. Perhaps the natural candidate for being the source of this creativity is the self.

The computational theory of mind has the mind as an essentially complex machine, but it fails. Can a more plausible theory be devised which reconciles both the metaphysical simplicity for which I have argued and the manifold complexity of the way the mind expresses itself? Attempts to answer this question are liable to drive one into what Russell somewhere described as 'soupy metaphysics' and I cannot venture too far into such territory here.<sup>2</sup> Some insight into how one might approach the problem can be gained by considering the 'unity in diversity' that is an essential feature of thought.

Peter Geach has argued that the 'activity of thinking cannot be assigned a position in the physical time-series' (1969: 34). His reason for this is that, though the expression of a thought using a sentence will be spread through ordinary time, one's grasp on the content must come as a whole. If it did not, then by the time one had reached '1066' in the sentence 'the battle of Hastings took place in 1066' one's consciousness of the other components of the thought would have passed into history. What the sentence expresses as a whole is the thought of which one is conscious. Something that has an essential unity finds expression in something that is complex. The position seems thus to be the following. The expression of a thought in a sentence is spread out in the normal 'flowing' empirical time. But the thinking of the thought which, in some sense, 'lies behind' (but not necessarily temporally before) this, is not temporally structured in the same way. Something which is implicit in the thought is laid out explicitly in the sentence. One experiences a thought *in* a sentence – or sometimes in other, non-verbal, images – but as a unity that a mere string of sounds or images does not possess.

Isn't this a somewhat mysterious doctrine? It is, but it is true to the phenomenology of thought. It can also be illustrated by appeal

---

<sup>2</sup> For a 'non-soupy' defence of simplicity rather different from mine, see Chisholm (1991).

to the distinction between two forms of potentiality. These forms of potentiality can be found in Aristotle, but also, I think, in our common conception of how things work. A hot object may be potentially cold, but to become cold is simply to change a property. 'Potentialities' in this sense signify the range of changes of which an object is capable. But when one exercises a specific ability – even more so, when one articulates a previously implicit thought – this is not a simple change, but a realization or externalization of something that was there but not 'laid out' in an explicit form. And Geach's point is that, when laid out or expressed it still retains the unity originally possessed, otherwise it could not be the expression of the content that it is.

Just as the complexity of a sentence or proposition expresses something which is, in a sense, a more primitive unity, perhaps the relation of the self to its various modes of expression is similar to this. In trying to make sense of this I shall draw on the neo-platonic account of the relation of the One and Intellect, and Aquinas on the Trinity – which are themselves, of course, connected.

#### V. PLOTINUS, ONE AND INTELLECT

Plotinus's metaphysics is based on the idea that intellectual complexity can unfold from the essentially simple, and he believes that this extends to the individual human subject.

One must, then, assume that a simple thing thinks itself, and investigate as far as possible how it does so ... (*Ennead* V.3. 1)

... we ought to think that this is how things are, that there is the One beyond being ... and next in order there is Being and Intellect, and the nature of Soul in the third place. And just as in nature there are these three of which we have spoken, so we ought to think they are present also in ourselves ... (*Ennead* V. 1. 10)

This naturally strikes the modern reader – especially the analytic one – as totally opaque. But, with the help of Geach's remarks concerning the unity of thought given above, perhaps we can entertain the idea that there may be a distinction between two kinds of simplicity. One, which we might call 'minimal simplicity', is the idea that something is simple when it possesses only one essential or internal simple feature. The other, which one might label 'undifferentiated simplicity' is when something possesses a nature which can only be expressed by attributing a complex set of properties but when those properties do not exist in the

thing as separately identifiable elements. This latter idea is, of course, a controversial one. Traditional theology talks of God's properties in this way, but the unity of thought seems similar to this, and, I am suggesting, the relation of the 'simple self' to its expression, in thought and action, is also analogous.

How are we to characterize the movement that is the expression of ourselves in thought and action? The CTM (which, if complete, would also be a Computer Theory of action, or, at least, volition) thought of this as a computational process, but we have rejected that. The inclusion of the semantic or meaning element introduces something which is, in a sense, informal, and could be designated "intelligence" or even "imagination" or "creativity", in a sense of those latter terms in which they differ little from intelligence construed in a wide ranging way. The account is "trinitarian" in that it gives the self three elements or moments: its essential simplicity, its expression as intellect and intelligent action and the process by which it moves from one to the other, namely imagination, creativity or intelligence.

It is not, I think, adequate to think of this as one thing – the self which is simple – *doing* other things, namely expressing itself through its intelligence. The interdependence of the elements is greater than this: they are the essence of the self. Plotinian language seems to me to be here appropriate. Talk of emanation or procession and return captures the informal intelligence by which the self expresses and 'unpacks' itself in thought and action and keeps returning to its implicit resources to develop and grow as a person handling himself in a world.

## VI. AQUINAS, THE TRINITY AND THE SELF

What are the similarities and differences in form between what I have claimed about the self and the western doctrine of the Trinity? First, the similarities. Both theories face the problem of how one can have an entity that is both simple and very rich (in God's case, infinitely so) in nature. Both involve a source paired off, so to speak, with something which constitutes its articulation or expression – the Intellect or Logos. And these two are related by a dynamic principle which is what enables one to be an expression of the other and both to communicate their nature.

Both theories claim that there is one substantial thing which has, in its own essential nature, three moments. The parallel between the two is that there is a sense in which one thing is equally present in its

metaphysically simple nature, in its plural, intelligent manifestations and in the process of unfolding the simple essence into this complexity.

But it is at this point that the differences become salient. It is a vital difference between the human self and the Divine Trinity that an individual human being is not three persons, so the presence of the self in its three moments does not give rise to three *hypostases*, three individuals. Given the similarities I have above claimed there to be, should it not be the case that we are, in some such way, triune?

The best way of approaching this problem is by trying to understand what Aquinas means by saying that the three persons of the Trinity exist by being *substantial relations*, for, by contrast, in our case the relations are not substantial in the sense of giving rise to three individuals. Exploring this difference can, I believe, help in explaining both this approach to the self and the meaning of Aquinas's account.

## VII. THE DIFFERENCE BETWEEN OURSELVES AND THE TRINITY

Aquinas does have a serious problem in explaining the notion of substantial relations. The persons have to be individuals, so they cannot simply be ordinary relations between the parts of one thing. On the other hand, they cannot be separate substantial things, for that would be tritheism. The Divine nature must be communicated to the three *hypostases* sufficiently fully to make them individual enough to each be designated 'persons', but not to such as to make them separate substances.

Aquinas claims that one substance – what is more, an absolutely simple one – can be three persons because of a relational structure internal to it.

For [the Persons] are many by reason of the distinction of subsistent relations, yet one God, nevertheless, by reason of the unity of subsistent essence. (SCG, IV, 14, 14)

The unity of the essence does not consist, as with people collectively, in the sharing of the same *kind* or *species*, in the manner of a universal with many instances, but in being the very same particular instance. The problem is with the role of the subsistent relations: how can relations within one substantial thing generate three individuals that can be truly categorized as persons?

This apparent contradiction makes sense if one takes as one's model the nature of the self, and this is what Aquinas does. He argues that, in thinking, something is generated which does not require that there be something which receives the thought, in contrast with what is the case

when someone produces a child, or a table, in which case there must be some matter that receives that which forms of the new entity. A thought is produced by a thinker, but it is not *made out of* something separate. The logic behind Aquinas's argument is, I think, this. If an object of kind *F* can generate something distinct from it of the same kind, but not by imprinting *F-ness* in some separate matter, then what is generated is not a different instance of *F* and so is substantially the same.

It might be helpful here to consider Aquinas's account of angels. Because they lack matter, each is a subsistent form and so they must all be of different species in order to be different individuals. This can prompt the thought that if (*per impossibile* in the case of angels) a subsistent form could replicate itself the result would be a complex which was, in a sense, two things and in a sense not.

The first worry that may arise with this approach is that it rests too much on the Aristotelian idea that it is matter that individuates, so that, if there is no matter involved, there cannot be two instances of the same form *qua* universal involved. This is, indeed, a very controversial principle, but I think we can defend Aquinas's overall position without relying on this principle, by emphasizing the analogy with thought. It is true that the thinker, or intellect, creates thoughts out of himself, or itself, but not out of any alien matter. But the problem will now be that the thought is not sufficiently 'other' from the thinker for it to be parallel to the case of a generated *person*. How can we combine the insight we can gain from the analogy with thought with the comparison with 'impossible twin angels'? The former helps explain the possibility of the generation of one thing from another without their being different substances, whilst the latter offers a model of the completeness of each individual so generated.

The difference between the Trinity and the human self becomes relevant here. There is a very limited sense in which we are a kind of trinity. There is a sense in which we are present in each of the three moments – in the simple self, in the outward expression of ourselves and in the process of moving from the one to the other. And we do have a kind of internal dialogue within ourselves between these moments.

I am present in the simple self, in the complex manifestations of my life and thinking and in the process from one to the other, but not wholly in any of them alone. But because we are finite creatures operating within time, our whole nature is never communicated between these *personae*, so to speak. I do show my nature in what I think and do, but only, so to

speak, in dribs and drabs. I think that it is reasonable to hypothesize that this is a feature of embodiment. We are dependent both on the brain and on the phenomenal realm, in the form of images and words, to work out and express our thoughts. This ensures that our whole nature cannot be articulated at once. (For an attempted account of embodiment which tries to explain these dependencies, see Robinson 1989.)

It is different in the case of God. What God shows forth in uttering the Word and in expressing His love (or will – the analogy varies) is always and eternally His full essence. So the very same individual essence has three complete expressions. Everything that is the Divine essence is eternally expressed in the Word and is contained in the Spirit that represents their dynamic link. This is why they are not simply *actions* of the Father, but forms of the Divine essence. In our case they are expressions of our nature, not communications of it whole.

### VIII. CONCLUSION

The account I have tried to develop can be taken on a variety of levels. Many mainstream philosophers, from Descartes to Chisholm, have defended the view that the self is a simple entity, and they therefore face the question of how this squares with the plurality of our mental features. I have tried to suggest how we might understand this. But, in the course of so doing I've drawn on the neo-Platonic account of the One and *Nous*, linked by the process of emanation and return. Finally, I compare this to the Christian doctrine of the Trinity. Our essential but finite and dependent simplicity images the absolute transcendent unity of the Father, our stumbling attempts at embodied rational self-expression parodies the perfect rational self understanding of the *logos*, and the creativity that ties them together in us palely images the role of the Holy Spirit. We are thus shadows and images of the Trinity as a whole, but, according to orthodox Christian doctrine, if these features in us are to be something better than a vain reflection of our Maker, they must somehow be incorporated in the their origin and source.<sup>3</sup>

---

<sup>3</sup> Acknowledgment: This paper was originally presented in Munich, at a conference for the Analytic Theology Project, generously funded by the John Templeton Foundation.

## BIBLIOGRAPHY

- Baker, M. 2011. 'Brains and Souls: Grammar and Speaking', in *The Soul Hypothesis*, eds. M. Baker and S. Goetz (London and New York: Continuum), pp. 73-93
- Bayne, T. and M. Montague, eds. 2011. *Cognitive Phenomenology* (Oxford: Oxford University Press)
- Chisholm, R. 1976. *Person and Object: A Metaphysical Study* (La Salle, IL.: Open Court)
- Chisholm, R. 1991. 'On the Simplicity of the Soul', *Philosophical Perspectives*, 5: 157-81
- Fodor, J. 2000. *The Mind Does Not Work That Way: The Scope and Limits of Computational Psychology* (Cambridge, Mass.: MIT Press)
- Geach, P. 1969. 'What Do We Think With?', in *God and the Soul* (Cambridge: Cambridge University Press), pp. 30-41
- Nagel, T. 2012. *Mind and Cosmos: Why the Materialist Neo-Darwinian Conception of Nature is Almost Certainly False* (New York: Oxford University Press)
- Robinson, H. 1989. 'A Dualist Account of Embodiment', in *The Case for Dualism*, ed. by J. R. Smythies and J. Beloff (Charlottesville: Virginia University Press), pp. 43-57
- Robinson, H. 2003. 'Dualism', *The Blackwell Guide to the Philosophy of Mind*, ed. by Stich and Warfield (Oxford: Blackwell), pp. 85-101. (Reprinted in *L'uomo a due dimensioni*, an Italian collection on dualism, ed. by Andrea Lavazz (Milan: Monadori, 2008), pp. 140-61)
- Robinson, H. 2003. 'Dualism', *The Stanford Encyclopaedia of Philosophy*, ed. by E. Zalta (Revised, 2007)
- Robinson, H. 2008-9. 'Vagueness, Realism, Language and Thought', in *Proceedings of the Aristotelian Society*, vol.CIX: 83-101
- Robinson, H. 2012. 'Substance Dualism and its Rationale', in *Free Will and Modern Science*, ed. by Richard Swinburne (Oxford: The British Academy and OUP), pp. 158-77
- Searle, J. 1980. 'Minds, Brains and Programs', *Behavioral and Brain Sciences*, 3: 417-57
- Williamson, T. 1994. *Vagueness* (London: Routledge).