The Thought Experimental Method: Avicenna’s Flying Man Argument*

ABSTRACT: No argument from the Arabic philosophical tradition has received more scholarly attention than Avicenna’s ‘flying man’ thought experiment, in which a human is created out of thin air and is able to grasp his existence without grasping that he has a body. This paper offers a new interpretation of the version of this thought experiment found at the end of the first chapter of Avicenna’s treatment of soul in the Healing. We argue that it needs to be understood in light of an epistemological theory set out elsewhere by Avicenna, which allows that all the constitutive properties of an essence will be clear to someone who understands and considers that essence. On our reading, this theory is put to work in the ‘flying man’: because the flying man would grasp that his own essence has existence without grasping that he has a body, connection to body cannot be constitutive of the essence.

KEYWORDS: Avicenna, self-awareness, dualism, essentialism, Islamic philosophy

No argument from the Arabic philosophical tradition has received more scholarly attention than Avicenna’s ‘flying man’ thought experiment. It has recently been hailed as a major contribution to the theory of self-awareness (Kaukua 2015: ch.2), and in the past it has been compared to Descartes’s cogito argument.¹ Though Avicenna alludes to the argument several times in his works (see the list in Hasse 2000: 81–82), the passage on which most scholarship has focused—and on which we will likewise concentrate here—is to be found on the last page of the first chapter of Avicenna’s (1959) treatment of soul in his Healing (al-Shifa’)—we will subsequently refer to this section of the Healing as On the Soul.² The Healing is

We are grateful for helpful suggestions from Salimeh Maghsoudlou, Helen Beebee, and Miranda Fricker as well as participants in a seminar on this paper at the University of Toronto. We are thankful to Dag N. Hasse and all other members of the Arabic reading group held at Würzburg and Munich, where we first started exploring the ideas that led to this paper. Finally we gladly acknowledge the DFG for support of our work under the aegis of the project ‘The Heirs of Avicenna: Philosophy in the Islamic East, 12th–13th Centuries’.

¹ This article is the second in a special series of commissioned articles on non-Western philosophies. The first article ‘Marxism and Buddhism: Not Such Strange Bedfellows’, by Graham Priest, appeared in Volume 4, Issue 1, pp. 2–13.

² For references and a critical analysis of this comparison see Druart (1988) and Hasnawi (1997). Sorabji (2004:168, 2006:222–26) draws a parallel between Avicenna’s flying man and the ‘the second use’ of Descartes’s and Augustine’s cogito (which is about proving the incorporeity of soul and thus fits well with our remarks in the conclusion to the present paper). Yet, Sorabji also emphasizes the necessity of self-awareness in Avicenna (which would correspond to the ‘first use of cogito’ in his language and which is less consonant with our reading).

³ We have used the Arabic edition by Rahman in Avicenna (1959), cited by page and line number.
distinguished from Avicenna’s other systematic philosophical compendia by its length and also by its self-conscious engagement with the Aristotelian tradition (for a detailed analysis of this aspect of the text in the case of the Metaphysics of the Healing see Bertolacci 2006). Our passage is a case in point. The first chapter offers a nuanced critique of previous attempts to define soul, with special reference to Aristotle’s definition that soul is, as Avicenna phrases it, ‘the first perfection of the natural, organic body to which it belongs to perform the actions of life’ (On the Soul 12.7–8; cf. Aristotle, De Anima 2.1.412b4–6).

The thought experiment needs to be understood in this context. It is an attempt to do what Aristotle’s definition failed to do in Avicenna’s opinion: determine whether it belongs to the soul’s essence that the soul be related to a body. By considering the flying man scenario, we come to see that the answer to this question is no. Thus, the upshot of the flying man thought experiment is akin to learning, for instance, that it is no part of the essence of a triangle that triangles should have three equal sides. A triangle may have three equal sides, but it need not, and likewise a soul may be related to a body, but it need not be. Several previous scholars have already appreciated this very specific and indeed rather narrow consequence of the flying man argument (at least as it is used in the chapter with which we are concerned), namely, that it excludes corporeality from the soul’s essence (Druart 1988; Hasse 2000; Sorabji 2006:222; Sebti 2000:59). Indeed, Avicenna could hardly be clearer that the thought experiment is supposed to tell us about the essence of the soul than he is in the opening of the famous passage, here quoted in full (On the Soul 15.17–16.17; we label the sections for ease of reference):

[A] We have now come to know the meaning of the name that applies to the thing called ‘soul’ through a relationship that it has. So it would be appropriate for us to occupy ourselves with grasping the essence (māḥryya) of this thing which is said to be ‘soul’ by the consideration [just discussed].

[B] We ought to indicate here an affirmation of the existence of soul which is an affirmation for us by way of admonition and calling to mind (ʿala sabīl al-tanbīh wa-l-tadhkīr), as an indication that will make a powerful impression on someone who has the capacity for noticing the truth by himself, without needing to be instructed, prodded, or turned away from sophistries.

[C] We say: one of us must suppose that he is created all at once, and created as perfect, but with his sight prevented from seeing anything external [to him]. He is created hovering in the air, or in a void, in such a way that the air does not buffet him so that he would have to feel it. His limbs are separated so that they do not meet or contact one another.

[D] Then, he considers whether he affirms the existence of his essence (dḥāṭ). He has no doubt in his affirmation that his essence is existent, even while he does not affirm any extremity among his limbs, nor
anything inward among his innards—not his heart or his brain—not anything external. Rather, he has affirmed his essence while not affirming for it any length, width, or depth. If in this situation he were able to imagine a hand or another limb, he would not imagine it as a part of his essence, nor as a condition for his essence.

[E] You know that what is affirmed is distinct from what is not affirmed, and what is acknowledged¹ is distinct from what is not acknowledged. Therefore, as to the essence whose existence he affirms, it is specific (khāṣṣiyya) for it⁴ that it is identical to him (anna-hā huwa bi-ʿaynīhī) and distinct from his body or his limbs, which he has not affirmed.

[F] Thus the alert person has a way to be admonished concerning the existence of the soul as something distinct from the body, or rather distinct from body (ghayr al-jism bal ghayr jism), and [a way] by which he may understand it and be aware of it. But if he is heedless, he may need to be prodded.

As we will see below, there is some dispute over the meaning of the word ḏhāṭ in [D] and [E], since ḏhāṭ can mean both ‘self’ and ‘essence’. But there can be no disagreement about māhiyya in [A], which is unambiguously Avicenna’s word for ‘essence’.

However, the very vividness and power of the thought experiment practically invites us to draw more dramatic and less technical conclusions, for instance, that one cannot doubt one’s own existence (hence the comparison to Descartes) or that the soul is self-aware (Marmura 1986:388). For Marmura this means that the argument begs the question because it needs to proceed from the assumption that the flying man would indeed be self-aware. But in fact this is a premise of the argument, not the conclusion, as we mention in what follows. This point has also been made by Kaukua (2015:34). Another problem is that the thought experiment as it stands seems wholly inadequate for the purpose of determining which essential properties the soul does or does not have. In this paper, however, we will argue that the argument can do exactly what Avicenna says it should do, so long as we make two admittedly controversial assumptions. First, that the flying man could indeed ‘affirm the existence of his essence’ [D]. Though Avicenna elsewhere insists that self-awareness is an ineliminable feature of human mental life, he makes no attempt to argue for that here (see Black 2008). Rather, he seems to proceed on the basis that the sufficiently insightful person invoked in [B] and [F] will simply have a strong intuition that the flying man does indeed know that his essence exists. Second, and far less obviously, Avicenna is assuming that someone who is in a position to ‘entertain mentally’ an essence can grasp all of the features

¹ There is a problem with the reading of the Arabic here; we agree with the one suggested in Hasse (2000:80): muqarr and yuqarra.

⁴ Reading la-hā, with several manuscripts. Hasse (2000:80) and Kaukua (2015:35) retain Rahman’s la-hu and take -hu to refer to the flying man.
that belong to that essence. If you can understand a triangle without considering a triangle as equilateral, then it is no part of the essence of triangle to have three equal sides. Likewise, so we contend, if the flying man can grasp the existence of his soul’s essence without grasping the existence of his body, then no relation to the body belongs to that essence.\(^5\) Avicenna’s epistemology, especially concerning the knowledge of essences, thus forms a vital background for grasping the function of the flying man.

1. Criticisms of the Flying Man

Avicenna’s thought experiment is a bit like Anselm’s ontological argument, in that there is general agreement that it is unconvincing, but general disagreement about where it goes wrong. Various diagnoses of Avicenna’s mistake have been offered, from which we will concentrate on two. Let us begin with the worry that the flying man argument involves an illegitimate form of inference. Just as Anselm has been accused of moving from the conceptual level to the level of reality in his proof of God,\(^6\) so Michael Marmura accuses Avicenna of executing ‘an unwarranted swerve from the hypothetical to the categorical’ (Marmura 1986:388). Marmura does not spell out what he means by this, but the thought is presumably that, in general, we cannot draw conclusions about what is in fact the case from counterfactual scenarios. This methodological principle is far from obvious though. After all, some modern-day philosophers think that the purported conceivability of (counterfactual) zombies can warrant conclusions about our (factual) minds (the touchstone of this modern debate is Chalmers 1996).

To this we can add two further observations about the way Avicenna sets up the thought experiment, which may help answer Marmura’s worry. For one thing, at least in this version, the flying man is not a hypothetical being whom we are to consider from a ‘third-person’ point of view, like a zombie. Rather, Avicenna says that ‘one of us (al-wāḥid min-nā)’ should imagine being in the scenario. Obviously the situation remains counterfactual, but Avicenna may be trying to get us to introspect about our own access to our essences and about what could and could not impede that access. If so, then he is using a counterfactual scenario to ‘direct attention’ to a fact about our real mental life. This could be why he labels the argument as a tanbīḥ, meaning ‘reminder’ or ‘admonition’, in other words, a prompt for reflection or suggestive argument rather than a full-blown proof.\(^7\)

For another thing, it is worth thinking about the sense in which the scenario is ‘hypothetical.’ Avicenna says in [C] that the flying man is ‘created,’ that is, by

\(^1\) The alert reader may note that in section [D] Avicenna speaks of the flying man affirming the existence of his own essence, not his soul’s essence. We will return to this issue below.

\(^6\) See Aquinas ST I Q.2, a.1, repl. obj.2, regarding Anselm’s ontological argument showing only that God has being in apprehensione intellectus and not in rerum natura.

\(^7\) Hasse (2000:87) also thinks the label of tanbīḥ can help resolve Marmura’s worry though he does not explain why. It seems that Hasse is thinking that the ‘logical status’ of tanbīḥ is weaker than Marmura presupposes: he cites Dimitri Gutas’s characterization of tanbīḥāt as mere ‘hints and guidelines’ rather than ‘ready-made arguments’. The thought would then seem to be that though you cannot prove something categorical on a hypothetical basis, you can be alerted to something categorical on such a basis. We will return below to this sort of ‘bar-lowering’ response.
God. Now, in Avicenna’s system God does not just immediately create human beings *ex nihilo*. Rather, divine causation is mediated by a series of subdivine cosmological principles. Thus, to say the least, it is rather unclear how God could summon the flying man into existence in the real world as Avicenna understands it. If we grant what seems to be presupposed by the thought experiment, namely, that God does in some sense have the power to create the flying man, then we must admit that the flying man is not merely *conceivable* but is actually *possible* within the causal structure of the real universe, in the good Avicennan sense that God could render him existent. This, incidentally, goes for other medieval thought experiments that invoke divine power when hypothesizing a counterfactual scenario, for instance, God’s creating a perfect sphere touching a perfect plane at only one point. Given divine power to do anything that is logically possible, everything conceivable becomes really possible (assuming, that is, that logical possibility marks the bounds of the conceivable). This is important because, as we will see later, Avicenna thinks the real possibility of an essence is the requirement for discovering the properties that belong to that essence. In other words, for him real possibility is enough to warrant categorical conclusions.

Let us now turn to another critique of the thought experiment, which has been considered by several authors, most recently by Jari Kaukua. He puts the worry this way: Avicenna ‘seems to commit the rather blatant fallacy of proceeding from an epistemic or phenomenological distinction to a metaphysical one’ (Kaukua 2015:37). In other words, the thought experiment trades on moving from an opaque to a transparent context. In [E], Avicenna says ‘you know that what is affirmed is distinct from what is not affirmed’. But from within an opaque context I can indeed affirm the existence of Cicero without having to affirm the existence of Tully. It is only within a transparent context that the truth values of the two affirmations must be the same. Applying this to the present case, we might suppose that the soul is *in fact* the body or some part of it. Then the flying man could certainly affirm that his soul exists without affirming that his body exists, but this is only because he does not, and indeed cannot, know that his soul is in fact his body.

This looks like a lethal objection to Avicenna’s thought experiment. Some interpreters have however suggested that the argument can be saved by ‘lowering the bar’ on Avicenna’s behalf. Emphasizing that this is an ‘admonition (tanbth)’ and not a proper demonstration, it has been suggested that the thought experiment is merely a kind of promissory note, offering us a ‘concrete idea’ of what the self ‘could’ be apart from the body, namely, a self-aware immaterial

---

8 For this idea proposed by Adam Wodeham, see Zupko (1993). Given Avicenna’s influence and the wide reception of his psychology, it is worth considering the hypothesis that the flying man argument itself spurred Latin scholastics to devise such thought experiments based on God’s power to actualize any possible scenario.

9 Kaukua also uses the word ‘opaque’ in explaining the objection. See also Sebti (2000:121): ‘to establish a distinction of reason between two realities is not a sufficient proof that they are really distinct’; and Black (2008:65): ‘This last move in the Flying Man, which is repeated in all of its versions, is of course problematic, since it seems to contain the obviously fallacious inference pattern, “If I know x but I do not know y, then x cannot be the same as y”’. 
Unfortunately, these face-saving proposals do not fit the text very well. The rule that ‘what is affirmed is distinct from what is not affirmed’ is stated confidently and with no caveat; if the conclusion were instead that these two things are possibly distinct, Avicenna could and should have said so. Nor is there a reference to a demonstration to be given later on that might show that they really are distinct. To the contrary, while Avicenna seems to grant that the thought experiment may not have its intended effect, this is only because it demands a high degree of perspicacity on the part of the reader. If the flying man does not persuade you, this is because you are a ‘heedless’ sort of person who needs ‘further instruction’ or ‘prodding’. The perspicacious reader, by contrast, will understand the point being made here about the soul’s essence.

2. The Context of the Flying Man

To understand what this point should be, we need to turn back to the preceding pages of the first chapter. As already mentioned above, these pages consist of a critical engagement with Aristotle’s definition of soul as the ‘perfection (kamāl)’ of the body. By this Aristotle would mean a form that supplies the body with a range of capacities ranging from the nutritive power to thinking (for a thorough discussion of the meaning of ‘perfection’ in the reception of Aristotle up to Avicenna, see Wisnovsky 2003:113–41). The problem is that this does not capture what the soul is in itself, but only gives us a grasp of the soul as extrinsically related to something else, namely, the body:

If we come to know that soul is a perfection, then—however we explain and elucidate ‘perfection’—we will not thereby come to know the soul in its essence (māhiyya) but only know it insofar as it is ‘soul,’ given that the name ‘soul’ applies to it not insofar as it is a substance but insofar as it governs bodies and is related to them. (On the Soul 10.15–18)

The same would apply to thinking of the soul as a ‘power (quwwa)’ for exercising certain activities. This designation ‘does not include reference to the essence of the soul as such, taken absolutely (dalāla ‘alā dhāt al-nafs min ḥaythu hiya nafs mutlaqan), but only in a certain respect’ (8.2–3).

Avicenna has been preparing the ground for these criticisms since the beginning of the chapter. There, he explained to us that when we use the word ‘soul (nafs),’ we are only talking about soul insofar as it is a causal principle for certain activities. ‘This expression,’ that is, ‘soul (nafs),’ ‘is a name for this thing not with regard to its substance (jawhar), but in respect of a relation (idāfa) it has, that is, in respect of its being a principle for those activities’ (4.10–12). To think of the soul like this is to grasp it in virtue of some accident that it has (4.14: min jihat ma-lahu ‘araḍ mā), rather than grasping its essence, something Avicenna compares to realizing that a moved thing has a mover, without knowing the essence (dhāt) of

\[10\] See Kaukua (2015:34); and Marmura (1986:393): ‘a thought experiment, not intended as a rigorous proof, will awaken them to this knowledge’, that is, ‘experiential knowledge of ourselves’.
the mover (5.2–3). As he also says early in the chapter, ‘on the basis of this accidental feature (al-‘arid) that belongs to it, we need to work on verifying its essence (dhāt), in order that we may make known its quiddity (māhiyya)’ (5.1–2). While this may sound as if Avicenna considers Aristotle’s approach deeply inadequate, that approach does suffice to prove that the soul exists because we know that there must be some principle that gives rise to activities in the body. Avicenna later on adds that it is from this perspective that we deal with the soul while doing natural philosophy, something that would otherwise be inexplicable given that the soul is an incorporeal entity and in this respect more germane to metaphysics (On the Soul 11.1–2; this issue has been investigated in Alpina 2016).

The structure of the chapter as a whole, then, is as follows. Avicenna establishes that there exists a principle we call soul, which can be grasped accidentally through the activities it makes manifest. This is what we do when we think of it as a power or perfection of the body. Grasping it in this way shows us that there is indeed a principle or entity for us to inquire about more deeply—again, think of the case of using a moved thing to establish the existence of a previously unknown mover. Aristotle’s definition merely does preparatory work, leaving unanswered the deeper question: What is the essence of this thing that we are calling ‘soul’? This, of course, is the question posed at the beginning of the flying man passage, which is worth restating here:

[A] We have now come to know the meaning of the name that applies to the thing called ‘soul’ through a relationship (idaya) that it has. So it would be appropriate for us to occupy ourselves with grasping the essence (māhiyya) of this thing which is said to be ‘soul’ by the consideration (iṭibār) [just discussed].

That is, we need to go beyond the ‘consideration’ or ‘point of view’ from which the soul is seen as a source of activities and find a way to grasp its essence: the task of the flying man thought experiment.

In light of this context, it seems undeniable that Dag Hasse is correct in saying that the word dhāt in the flying man passage means ‘essence’, and not ‘self’ (Hasse 2000:83). To support this, Hasse refers to another use of the thought experiment in chapter 5.7 of the On the Soul, where the flying man grasps his anniyya (‘core being’). He rightly alludes to the context here in chapter 1.1 as well. To his persuasive argument, we would add that the word dhāt has appeared throughout the chapter and consistently meant ‘essence’, being used interchangeably with māhiyya. In addition to the uses of dhāt in the passages just quoted, we find in this chapter an explanation that being a substance requires never being in something as in a subject, the way that an accidental property exists in a substance. Such subject-independence must belong to the substance ‘in its essence (dhāt)’, something that can be verified in the following way: ‘If you consider and inquire into its essence (idhā tā’amalta dhātahu wa-nazarta ilay-hā), and then no
subject is found for it at all, it is in itself substance’ (9.15–16; for the logical background of this argument see Benevich 2017).

None of this is to say that the flying man argument articulates a complete, scientific understanding of the soul. Rather, it has the limited and in fact purely negative role of showing us that whatever the essence of soul is, it does not include the ‘accident’ of relation to the body that formed the basis of the Aristotelian definition of the soul as a ‘perfection of the body’. Again, Avicenna states this fairly explicitly, when he concludes the thought experiment in [F]: ‘The alert person has a way to be admonished concerning the existence of the soul as something distinct from the body, or rather distinct from body (ghayr al-jism bal ghayr jism)’. The latter contrast probably means that my soul is distinct not just from my body, but from body in general. We should also understand this to be the purport of his remark in [B], which labels the flying man argument as a way to ‘affirm the existence of the soul (ithbāt wujūd al-nafs)’. This formulation is misleading because it makes the thought experiment sound as though it establishes whether the soul exists at all, which has no doubt encouraged the comparisons to Descartes’s cogito. But as we have seen, the rest of the chapter has already shown the soul exists, on the basis that there must be some principle for the activities we see exercised in the body. Rather, Avicenna here uses ‘existence of the soul’ as shorthand for ‘existence of the soul as distinct from body’, as he spells it out fully in [F]. (We are here in agreement with Alpina 2016:191.)

We are now in a position to articulate the purpose of the flying man thought experiment. It is intended to give us a very particular insight about the essence of soul, namely, that this essence requires no connection or relation to body—a view that might have been mistakenly inferred from Aristotle’s way of describing the soul as the ‘perfection of the body’. To put it more positively, the flying man thought experiment helps us to see that it is essential to soul that it be ontologically independent of body. The question now is, how can the thought experiment do that? A clue has been given earlier in the chapter, where Avicenna commented on the limitations of the Aristotelian approach:

We must separately devote a different investigation to the essence (dhāt) of the soul; if we had thereby come to know the essence of the soul, then we would have had no difficulty about which category it falls into. For whoever knows and understands the essence of a thing, and then turns his own attention (ʿaraḍa ʿalā nafsīhi) to the nature of some essential feature (tablāt amr dhātī) that belongs to [that thing], has no difficulty concerning the existence of [that feature] for [that thing], as we have explained in the logic. (11.3–7)

12 An attentive reader of the On the Soul I.1 may notice that at 5.3–20, Avicenna already provides an argument that ‘the essence of soul is not a body’. However, this does not contradict our interpretation that the flying man serves to establish the incorporeality of soul. For the argument at 5.3–20 is intended to show that soul is not just a body, which leaves open the possibility that soul still has some essential relation to body, as the Aristotelian definition would suppose. One may compare the difference between these two arguments to the difference between Plotinus’s arguments against Stoic materialist psychology, on the one hand, and Aristotle’s on the other hand (Enneads IV.7.5 and IV.7.8 respectively).
This passage is crucial for understanding the flying man argument and evokes Avicenna’s confidence that someone who grasps a given essence will be able to discern the features of that essence. For instance, someone who grasps what a triangle is will be able to discern that triangles always have three sides, or that they are a species of geometrical figure. This is much like the formulation used for the subject-independence of substance, where we were told to ‘consider and inquire into its essence’. To understand this fully, we need to do as Avicenna suggests and look back to his explanation ‘in the logic’.

3. Grasping Essences

We have just argued that the crucial distinction for the argument in On the Soul 1.1 is that between essential and accidental attributes. Aristotle has offered a useful account of soul, but one limited to grasping soul through its accidental relation to body, and the flying man argument seeks to improve on this by giving us an insight into the soul’s essence. In his logical works, Avicenna offers various ways of understanding the difference between essential and accidental features, but the one that is relevant for us is the definition of essentiality and accidentality presented in his Introduction to his logical works (Avicenna 1952, referred to in what follows as Introduction). This part corresponds to Porphyry’s famous Isagoge, which prefaced the Peripatetic logical corpus in the curriculum of study of late antiquity and the Middle Ages. Here, right at the beginning of his analysis, Avicenna states that everything in this world has an essence (dhat) or a quiddity (mahiyya) through which it is what it is (Introduction 28.13). Some attributes belong to this essence; others are accidental.

Essential attributes are features whose composition is the necessary and sufficient condition for the establishment of an essence. For instance, human is defined as rational animal, and thus both animality and rationality constitute the definition of human and are essential for him. These attributes are in this sense intrinsic to human. The constituents of these attributes will also constitute the essence. Given that animal is defined as ‘a corporeal substance that possesses sense perception and moves by will’, the attributes mentioned in this definition will also be among the essential attributes of human. The same applies to the definition of each of these attributes, such as corporeal substance, until one exhausts all attributes. ‘When they are joined together, one essence comes to be from their composition, which is the essence of human’ (29.6–7).

By contrast, accidental attributes are neither sufficient nor necessary for the formation of the essence to which they pertain. These accidental attributes might be specific and even extrinsically necessary, but they are not essential (28.15–30.6). To use the classic example, it is a necessary consequence of being a human that one is able to laugh, but being able to laugh is not part of what it means to be a human. To be a human is to have the relevant constitutive properties, such as animality and rationality. Avicenna thus offers a conditional test for the difference between the essential and the accidental: if Y is required for the constitution of X, Y is essential for X (here and in what follows Y will stand for a predicate of X, not for a separate cause, which might be necessary for its effect in a different
sense). Everything else, even if it necessarily pertains to X, is extrinsic and accidental for it. Avicenna sums up the position as follows: ‘a “thing” has a real essence (dhātan ḥaḍiqiyyatan), and attributes pertain to it. The reality of [that] thing is composed out of some of them, whereas others are accidents that are not necessary for its existence, while still others are accidents that are necessary for it’ (30.7–9).

Avicenna furthermore considers that two sorts of existence may attach to essence, ‘mental’ and ‘concrete’. The latter is commonsense extramental existence, as when the essence of human is instantiated in Socrates. The other is an essence that is grasped by us so as to exist in the mind as a universal concept. Crucially, in both cases the essence may be qualitatively one and the same: the essence of human instantiated in Socrates fully corresponds to the idea of human we think about. (This is the normal situation: in the case of fictional or noninstantiated entities, the essence has only mental existence.) As Jon McGinnis has rightly suggested, the idea of the ‘human qua human’ and the strict correspondence between mental forms and extramental entities provides Avicenna with a strong epistemological basis for claiming that we can and do have knowledge about essences outside our minds (McGinnis 2007a:170).

Let us now combine the two points just discussed: Avicenna’s compositional theory of essences and his account of mental existence. The outcome is that essences, as they are constituted by the aggregate of their intrinsic attributes in extramental reality, can be conceptually grasped by our minds only once the same aggregate is present to us. That is, if animal, rational, corporeal substance, and the rest compose the essence of human in Socrates outside the mind, then these attributes must all be contained in the essence that is in the mind too. Avicenna draws this conclusion explicitly in his Introduction 1.6:

If a quiddity has constituents that precede it insofar as it is that quiddity, then the quiddity does not occur (tahṣulu) without their preceding it, and when a quiddity fails to occur, it occurs neither as an object of the mind (maʿqūl) nor as a concrete object (ʿayn). Therefore, when [the quiddity] arises as an object of the mind, it occurs only so long as that which constitutes it occurs in the mind together with it, in the same way as it constitutes it. Hence, if this occurs in the mind, it cannot be negated. So these constitutive attributes must be grasped mentally together with the conception (tasawwur) of the thing, so that their existing for it cannot be unknown (la yūjhalu wa yūjūdūhā la-hu) and it is impossible to negate them of it, as if the quiddity could be established in thought (dhihn) even as [these attributes] are actually eliminated from it. By their ‘occurring in the intellect’, I do not mean that they are actually brought to mind (bāl), since many objects of the mind fail to come to attention. Rather, I mean that when one does bring [a quiddity] to attention and at the same time brings to attention what constitutes it, so that both are brought to mind in actuality (mukhtar bi-l-bāl bi-l-fīl), then [the constituting attributes] cannot be negated of it, as though you could find the quiddity to be actually stripped of these [attributes] while it is conceived, I mean, when the quiddity is
conceived in thought. If this is so, the attributes which we called ‘essential (dhāṭī)’ for mentally grasped concepts (li-l-ma’anā al-ma’qūla), must necessarily be grasped for the thing in this way, since the quiddity cannot be conceived of in thought, unless their conception precedes it. (34.13–35.5)

In this passage, Avicenna adopts a biconditional criterion for essential attributes:

\[ Y \text{ is essential for } X \text{ if and only if } X \text{ cannot be grasped without grasping } Y. \]

Or to put the same point negatively: If \( X \) can be conceptualized without \( Y \), then \( Y \) is an accidental feature of \( X \). This may seem counterintuitive. Certainly, I do not seem to think actively about \textit{animal} whenever I deploy my grasp of the essence of \textit{human}. Avicenna anticipates this objection. He clarifies that he does not mean explicit apprehension of both concepts together. Rather, his idea is that if one does grasp \textit{human}, then as soon as one explicitly reflects on (‘brings actively to attention’) \textit{animal} and \textit{human} simultaneously and asks oneself how these two concepts relate, one will see that \textit{animal} cannot be negated of \textit{human}.

It will be important for our discussion of the flying man thought experiment that Avicenna phrases his criterion in terms of existence conditions. Thus, he says in the passage just cited that the essence does not ‘occur’ without the constituents occurring and also that someone who is conceptualizing an essence cannot fail to know that the constituents ‘exist’ for it. As Avicenna says:

\[ \text{If both things are brought to your attention altogether, you cannot negate that which is a constituent from that which is constituted by it, in such a way that the constituted could exist in its quiddity in the mind while that which constitutes it would not exist [in the mind].} \]

\textit{(Introduction 35.12–4)}

To use the same example as above, we can say that since \textit{animal} is a constituent of \textit{human}, nothing can exist as a human without existing as an animal. More generally, we can state a second biconditional rule:

\[ Y \text{ is essential for } X \text{ if and only if } X \text{’s existence presupposes } Y \text{’s existence.} \]

13 Notice that the biconditional would not hold if we said ‘\( Y \) is necessary for \( X \)’ instead of ‘\( Y \) is essential for \( X \)’ because one may fail to grasp necessary accidents, called ‘concomitants (lawāzīm)’, while grasping an essence. For instance, a conceptualization of \textit{human} will not immediately reveal that humans are all able to laugh. Often great effort may be needed to discover necessary accidents: consider a mathematical case like the Pythagorean theorem, which necessarily applies to right triangles but is not part of the relevant essence and is thus not immediately evident upon conceptualization of right triangle.

14 Again, this is not true for necessary accidents, or ‘concomitants’ (cf. the previous note). They depend on and presuppose the essence, much as the essence presupposes the essential properties. Thus, for instance, \textit{human} is a precondition for being capable of laughing, not vice versa, just as \textit{rationality} is a precondition for \textit{human}, not vice versa.
Again, we should not interpret this rule as if it applied only to concrete, extramental things. As Avicenna explicitly says, he is talking about essences insofar as they arise either concretely or in the mind. Thus, we can reformulate the rule just given as follows:

\[ Y \text{ is essential for } X \text{ if and only if } X's \text{ mental existence presupposes } Y's \text{ mental existence or } X's \text{ concrete existence presupposes } Y's \text{ concrete existence.} \]

We will henceforth call this sort of reflection on attributes Avicenna’s ‘conceptual test’: in order to grasp an essence we must have grasped all its essential attributes so that we will have no doubt as to whether this or that attribute belongs to it essentially or accidentally. Thus, the fact that human can be grasped without having grasped pale or capable of laughter shows that the latter properties are not essential to human.

One might argue that this ‘conceptual test’ is a deeply problematic procedure. How could I ever be sure that I have formed the concept of human that fully corresponds to the extramental instantiations of the essence human? I might worry that rationality is not really essential for human and that it only seems to be ineliminable because I have formed the wrong concept, or I might unwittingly have failed to grasp some of the constituent attributes so that my grasp of human is still incomplete. These questions raise serious problems for Avicenna’s theory, as they would for many realist epistemologies. We cannot address them here and will simply mention that Avicenna has a developed theory of concept formation involving both abstraction from particulars and some sort of assistance from a cosmological intellectual principle, which could help him solve these issues (on concept formation in Avicenna, see Hasse 2001 and 2013, McGinnis 2007b, D’Ancona 2008).

Avicenna’s ‘conceptual test’ finds several applications throughout his philosophical works. For instance, it is important in proving the central idea of his metaphysics: the essence-existence distinction. He argues that we can conceive of the essence of human or triangle without knowing whether they exist, which means that existence is accidental to them (Avicenna 2008:155.1–3). Another example would be Avicenna’s proof that finitude is not essential for bodies, again, because we can conceive of corporeality without presupposing finitude (Avicenna 1985:498.14–19). And as we will now show, the flying man argument is in fact a further application of the conceptual test.

4. Back to the Flying Man

Avicenna makes use of his conceptual test twice in On the Soul 1.1, indeed in passages we have already cited above. In the first case, it supports his rejection of

---

15 Elsewhere, Avicenna similarly proves that motion is not essential for bodies. Taneli Kukkonen noted this example in his analysis of thought experiments (Kukkonen 2014:435). The thought experiments Kukkonen discusses in his paper are in fact conceptual tests of essentiality and should be read in light of the epistemology we have outlined in this paper.
the Aristotelian definition of soul as ‘the perfection of body’. As we saw, he says in that context that ‘whoever knows and understands the essence of a thing, and then turns his own attention (ʿaraḍa ʿalā nafsihi) to the nature of some essential feature (tabʿat amr dhāṭi) that belongs to [that thing], has no difficulty concerning the existence of [that feature] for [that thing]’. His point in the passage is to show that Aristotle’s definition is not up to the challenge of determining that the soul is a substance, a conclusion that will ultimately be established only in On the Soul 1.3 (cf. Hasse 2000:86). The argument can thus be analyzed as follows:

1. An essential definition provides knowledge of all essential attributes of an essence.
2. Belonging to a category is an essential attribute of essence.
3. Defining soul as the ‘perfection of the body’ does not provide us with knowledge as to which category the soul belongs to.

Conclusion: ‘Perfection of the body’ is not an essential definition of the soul.

Avicenna borrows his first two premises from his works on logic, as he explicitly acknowledges. Premise 1 is precisely the epistemological rule we observed in the previous section: knowing an essence presupposes knowing all its essential attributes. Premise 2 comes from his Categories in which Avicenna argues that being a substance is an essential feature of substances (see Benevich 2017). To this Avicenna only needs to add premise 3, which follows from his previous discussion on the Aristotelian definition of soul, to reach the desired conclusion that Aristotle’s definition fails to reveal soul’s essence. Again, this is not to say that soul is not the perfection of the body. Rather, calling soul the perfection of the body merely picks out its relation to bodies, something affirmed and studied in natural philosophy.

Having argued against the Aristotelian definition of soul, Avicenna frees himself from one definition of soul that explicitly binds soul to body. Yet, he has not ruled out the general idea that the soul may be somehow essentially related to the body. This is the purpose of the flying man argument, which deploys the conceptual test a second time. After announcing in section [A] that the argument will tell us something about the soul’s essence (māḥiyya), he proceeds to give us a thought experiment showing that someone can grasp the soul without even having the concept of a body. This shows that no corporeal ideas can be involved in the essence of soul; to put it another way, the definition of soul would not even need to mention the body. Or to put it yet another way, in terms of the criterion given above:

\[ Y \text{ is essential for } X \text{ if and only if } X \text{ cannot be grasped without grasping } Y. \]

We can substitute ‘a connection to body’ for \( Y \) and ‘soul’ for \( X \):

A connection to body is essential for soul if and only if soul cannot be grasped without grasping a connection to body.
But grasping the soul without such a connection is precisely what the flying man in the thought experiment does. We may conclude that although a human soul, such as the flying man’s, does indeed have a connection to the body, this connection is accidental.\textsuperscript{16}

Let us now return to the objections that have been raised against Avicenna’s thought experiment. First, the purported shift from a transparent to an opaque context. We already mentioned some possible responses above, but the best answer to the objection should now be clear. If we interpret the flying man argument against the background of Avicenna’s epistemology, in particular his conceptual test for essentiality, the argument is perfectly valid. As we can judge that capacity to laugh is a mere accident of human because we can conceive of human without ascribing capacity to laugh to this essence, we can also judge that a connection to body does not feature in the essence of soul because we can conceive of soul without a connection to body. In response, one might argue that our interpretation of the flying man just shifts the problem. In fact, it turns out that Avicenna’s whole epistemology fails to distinguish between the opaque and the transparent. This is, in effect, the worry we already mentioned above: how can we judge the constituents of the mind-independent essence of human just relying on an analysis of the mind-dependent concept of human? Yet, we have also seen that Avicenna has an answer to this worry. Human as existing in the mind must fully correspond to the extramental essence of human in Socrates due to the fact that both are instances of one and the same quiddity or essence, which in itself is neither universal nor particular, neither mental nor concretely instantiated. Thus, Avicenna does not so much overlook the distinction between the opaque and the transparent as give us grounds for rejecting it, at least in the context of grasping essences. To be trapped in an opaque context, unable to conceptualize an essence in a way that corresponds to the essence’s concrete instantiation, would for him mean failing to grasp the essence at all. The upshot is that pressing this objection against Avicenna’s flying man argument would require critiquing his essentialist metaphysics and epistemology as a whole.

Another objection we considered above was that the thought experiment illicitly shifts from a hypothetical situation to a categorical conclusion. We noted above that for Avicenna, considering a merely possible scenario can warrant such conclusions, and now we are in a better position to see why. For Avicenna, if an attribute $Y$ is essential to some $X$, then mental inspection will show that $Y$ cannot fail to belong to $X$ under any circumstances. Conversely, we need only find one (possible) circumstance in which $X$ lacks $Y$ to secure the conclusion that $Y$ is not essential to $X$. In contemporary parlance, we might say that if $Y$ is essential to $X$, then there is

\textsuperscript{16} It should be noted that the conceptual test envisioned in Avicenna’s Introduction only lets us know whether $Y$ is a constituent of $X$ or not. It does not tell us whether $Y$ is necessarily connected to $X$ as an accident, as having the sum of internal angles equal to two right angles is a per se accident of triangle. To establish what is and is not a per se accident, one should look for a middle term, which will not necessarily be obvious upon mental inspection of the essence. Hence, the flying man argument (on our reading at least) does not rule out that a connection to body is a necessary accident of soul, only that such a connection is part of the soul’s essence and hence its definition. Avicenna offers the additional proof in On the Soul 5.2 and 5.4, where he uses the capacity to conceive of universals as a middle term. On this argument see Adamson (2004).
no possible world in which $X$ exists without $Y$. Of course, Avicenna does not operate with the notion of possible worlds, but he does operate with the notion of mental existence. Where we would speak of a ‘counterfactual thought experiment’, Avicenna would speak of something that has only mental existence; indeed, we might say that a thought experiment gives us access to something that is factual, since being in the mind is a way of existing. But as we have seen, the conceptual test of essentiality applies to mentally existent things just as well as to concretely existent things, because the essence as such is the same whether a thing exists in the mind or in the world. Recall our third formulation of Avicenna’s biconditional criterion for essentiality:

$$Y \text{ is essential for } X \text{ if and only if } X\text{'s mental existence presupposes } Y\text{'s mental existence or } X\text{'s concrete existence presupposes } Y\text{'s concrete existence.}$$

A person in the situation of the flying man has his or her soul as an object of mental grasping (it is $ma'qūl$) and thus as mentally existent while his or her body is not mentally existent. This shows that the body is not essential to the soul. It is in this way that we can learn about the essential properties of soul from the ‘hypothetical’ flying man thought experiment.

5. The Role of Self-awareness

Though our reading of the thought experiment shows the two earlier objections to be unfounded, it may seem to open Avicenna to a new objection: How can we be so confident of discerning which properties the soul does and does not have essentially? We are still only in the opening chapter of the psychological section of the Healing and as yet have no proper definition of the soul to work with (for instance, as mentioned above, it has not yet been established that the soul is a substance). Admittedly, the flying man argument does not establish any positive attribute as being essential to the soul, but only rules out a candidate attribute, namely, connection to the body. On our reading, though, we are able to rule out this attribute only because we have a ‘conceptualization’ of the soul in a situation where the body is not being ‘conceptualized’. Why suppose that the flying man is able to conceptualize his soul?

It is here that it becomes relevant to invoke the phenomenon most readers take to be central to the thought experiment: self-awareness. As Avicenna says elsewhere, we are constantly aware of ourselves, even when asleep (see, e.g., Kaukua [2015:80]). His idea then seems to be that self-awareness gives us sufficient access to the soul and its existence conditions that we are able to perform the conceptual test involved in the flying man. Note that on our interpretation, the flying man thought experiment is not primarily about the ‘self’. As we argued above, $dhat$ in section [D] of the thought experiment should be translated as ‘essence’, and the framing of the whole passage, in sections [A] and [F], shows that Avicenna is offering us an argument that concerns the soul. Self-awareness comes into the picture only because it is our capacity for such awareness that guarantees the
soul’s conceptualization. It should be noted however that readers not convinced of our translation could still accept the rest of our interpretation. If dhāt in [D] means ‘self’ and not ‘essence’, then Avicenna simply invokes the phenomenon of self-awareness to show us a way to bring out conceptualization of the essence of soul to full attention. Still, we do prefer to understand dhāt as ‘essence’ because it makes clearer how the thought experiment could do what Avicenna promises in [A], by giving insight into a quiddity (māhiyya).

One may worry here that there is a slip from saying that the flying man is aware of his own ‘essence’ in [D] to claiming that the flying man is aware of, or conceptualizes, the essence of his soul. To which we would reply that this is a distinction without a difference: the flying man just is his soul. This must be the case, since otherwise the connection of his incorporeal soul or self to the body could not be accidental and his personal identity would not continue after the death of his body. Avicenna may be making this point explicitly in the rather inscrutable comment found in [E], which says that ‘the essence whose existence he affirms’—that is, on our reading, the essence of ‘this thing which is said to be “soul”’ (see [A])—‘is identical to him’.

As is clear from the idea that we are self-aware even in sleep, self-awareness is not always, or even usually, something to which we actively attend. Therefore, the immediate conceptualization of the soul that self-awareness makes possible will usually remain tacit. It is perhaps for this reason that Avicenna refers to the argument not as a ‘demonstration (burhān)’ but as an ‘admonition (tanbih)’. Avicenna is trying to draw our attention to something we already know, if only tacitly, namely, that we are always aware of our soul in a way that does not involve being aware of the body. If we cast our mind back to the passage from his Introduction, we will recall that he there admits that the connection between constituent attributes and a quiddity may not ‘actively come to mind (khatāra bi-l-bāl bi-l-fīl)’ even though that connection is tacitly something we know. The purpose of the flying man thought experiment, then, is to ‘remind’ us that we already have a conceptualization of our own souls, which is enough to give us access to the existence conditions of soul—in this case, that a connection to body is not an existence condition. This is also why Avicenna remarks in [B] and [F] that the reader will need to be sufficiently perspicacious to benefit from the flying man argument: Avicenna can help by offering a vivid thought experiment, but the rest is up to us.

Note that this is not a case of ‘lowering the bar’, with the argument being a ‘mere’ admonition or reminder as opposed to a demonstration. To the contrary, if Avicenna thinks that we have immediate access to our souls, then the resulting conceptualization has, if anything, a stronger epistemic status than a demonstration. For Avicenna demonstrations primarily derive necessary accidental properties from an essence, for example, by showing that the ability to laugh is implied by the essence of human or that the essence of triangle entails that a triangle’s internal angles are equal to two right angles. The flying man argument is not like that. It does not establish an extrinsic necessary property. Rather, it invokes our intimate grasp of our own souls to show us something about the soul’s intrinsic essential features, namely, that these features do not include a need for the body.
6. Conclusion

We began this paper by noting that some interpreters have seen a parallel between the flying man and Descartes’s cogito. We close by noting that on our interpretation, there is actually a closer similarity between Avicenna’s thought experiment and a different passage found in Descartes. In the sixth Meditation, Descartes argues that ‘the fact that I can clearly and distinctly understand one thing apart from another is enough to make me certain that the two things are distinct, since they are capable of being separated, at least by God’. From this he concludes: ‘Thus, simply by knowing that I exist and seeing at the same time that absolutely nothing else belongs to my nature or essence except that I am a thinking thing, I can infer correctly that my essence consists solely in the fact that I am a thinking thing. . . . Accordingly, it is certain that I am really distinct from my body, and can exist without it’ (Descartes 1984: 54; this parallel has already been noted by Druart 1988: 34–35). Here, Descartes infers from the conceivability of himself as a ‘thinking thing’ without body that he is distinct from his body; he even puts the point in terms of existence conditions. Thus, he deploys something very like Avicenna’s conceptual test for the essentiality of attributes. Both ask whether body is essential for soul and deny it, given that one can conceive of body without soul. This is not to deny that there are differences between the two thinkers. For Descartes the test’s validity rests upon the claim that God could separate two things without destroying them even if they are always found together. For Avicenna, it rests upon a deep confidence in the power of the human mind to see what is, and is not, essential.

PETER ADAMSON AND FEDOR BENEVICH
LMU MUNICH
peter.adamson@lrz.uni-muenchen.de
benevich@lrz.uni-muenchen.de

References


