Fundamentality and the Self

Benjamin Brast-McKie[*]

April 11, 2024

Abstract

Some of the most important philosophical and religious claims cannot be articulated in first-order languages, motivating the development and use of languages with greater expressive power. This paper investigates the nature of the self. In particular, I will seek to regiment the Upaniṣadic claim, '*sa eṣa neti netyātmā*' (`NA`), which Olivelle (1998, p. 101) translates as, 'About this self (*ātman*), one can only say 'not—, not—'.' After presenting the context of the Śākalya Dialogue from the *Bṛhadāraṇyaka Upaniṣad* in which this claim occurs, I will regiment `NA` as follows:

(NA) For any way of being, being that way does not strictly explain what it is to be the self. (In symbols, $\forall X[X \nleq \lambda x.(\alpha = x)]$.)

It follows from NA that the self is *fundamental* since this is what it is to not have any strict explanation. I will conclude by examining the relationship between the fundamentality of the self and the nature of the Absolute (*brahman*). In particular, given the assumptions that every way of being is to be weakly explained by the way that the Absolute is, and that the self is the Absolute, we may show that the self is fundamental.

1 Introduction

The Yājñavalkyakāṇḍa (chapters 3-4) of the *Bṛhadāraṇyaka Upaniṣad* (BU) begins with a competition proposed by King Janaka, where the most learned of the Brahmin priests will win a thousand cows with ten pieces of gold tied to the horns of each. Confident in his prowess, Yājñavalkya claims the prize, telling his pupil to drive the cows away before defending his superiority by answering the questions of the other Brahmin priests in attendance. Throughout these competitive exchanges, a number of interesting and enigmatic claims are made, including the following claim from the Śākalya Dialogue (SD):

(`NA`) About this self (*ātman*), one can only say 'not—, not—'.[1]
sa eṣa neti netyātmā. (BU 3.9.26)

[*] `benbrastmckie@gmail.com`

[1] All translations will be drawn from Olivelle (1998).

The use of 'not—, not—' (henceforth *néti néti*) is explicitly specified as the method of indication (*ādéśá*) in the Gārgya Dialogue (GD) preceding SD.[2] Whereas *néti néti* was traditionally translated as either 'not so, not so' or simply 'not, not', Geldner (1908, 1928) interpreted *néti néti* as a double negation, where this view was later championed by Slaje (2010, p. 14).[3] In opposition to this approach, Acharya (2013) presents compelling evidence against Geldner's reading, arguing that the aim of the broader discourse of BU is to, "highlight the importance of the method of rejecting all fixed definitions" (p. 29). Drawing on the formal methods of modern logic, this paper sharpens this reading by regimenting `NA` in the context of SD as a second-order claim whose quantifiers bind variables in predicate rather than nominal position.

Even though higher-order logics were unknown to the authors of BU, the axiomatic method provides an excellent alternative to the attempt to provide explicit definitions of the various concepts that they used. Rather, we may describe the relationships that a select range of concepts bear to each other by studying the roles that they play in a given context. By adjusting the axioms and rules of inference included in a logic for these concepts we may be explicit about what roles each of these concepts are assumed to play as well as their interaction with each other. In addition to specifying such a logic, I will explore the logical relationships that hold between the following claims:

(NA) For any way of being, being that way does not strictly explain what it is to be the self (i.e., $\forall X[X \nless \lambda x.(\alpha = x)]$).

(BF) For any way of being, being that way is weakly explained by the way the Absolute (*brahman*) is (i.e., $\forall X[\lambda x.(\beta = x) \leqslant X]$).

(AB) The self (*ātman*) is identical to the Absolute (*brahman*) (i.e., $\alpha = \beta$).

I will argue that these principles regiment claims that occur throughout BU. Given minimal assumptions about the strength of the background logic, I will derive NA from BF and AB. In addition to shedding light on the logical relationships that hold between some of the principle claims made in BU, this derivation provides evidence in support of the traditional reading of `NA` that Acharya (2013) defends. Accordingly, this paper aims to demonstrate the supporting role that formal logic may play in textual analysis.

Before presenting SD in which NA first occurs, it will be important to examine the earliest occurrence of *néti néti* in BU. Accordingly, §2 will review Acharya's (2013) reading of *néti néti* as it occurs in GD. After identifying a textual basis for BF in §3, I will argue in §4 that NA provides a natural regimentation of `NA`. I will then demonstrate in §5 that only a few natural rules of inferences are required to derive NA from BF and AB, where AB may be found to occur in various guises throughout BU. After defending each of these natural rules of inference in §6, I will conclude by reflecting on the broader role that formal logic may play in interpreting the *Upaniṣads* in §7.

[2] See Acharya (2013, p. 20) and Acharya (2017) for this reading of *ādéśá*.

[3] See Acharya (2013, p. 5) for citations and relevant discussion.

2 Gārgya Dialogue

The second chapter of BU begins with an offering from a learned Gārgya to tell, "a formulation of truth (*brahman*)," to king Ajātaśatru of Kāśi. After venerating the sun, moon, lightning, space, wind, fire, waters all as *brahman*, Gārgya turns to venerate the person in a mirror, the sound drifting behind a man as he walks, the person here in the quarters, the person here consisting of shadow, and the person here in the body (*ātman*) all as *brahman*.

In addition to rejecting the adequacy of each of Gārgya's specifications of *brahman*, Ajātaśatru asks, "Is that all?" When Gārgya does not answer, the roles reverse, where Ajātaśatru asserts, "It isn't known with just that," prompting Gārgya to ask, "Let me come to you as your pupil." Ajātaśatru responds by taking him to a sleeping man, rousing him from his slumbers, and asking "When this man was asleep here, where was the person consisting of perception? And from where did he return?" After Gārgya admits that he does not know, Ajātaśatru replies, "When this man was asleep here, the person consisting of perception, having gathered the cognitive power of these vital functions (*prāṇa*) into his own cognitive power, was resting in the space within the heart." By taking the vital functions with him, Ajātaśatru explains that, "[w]herever he may travel in his dream, those regions become his worlds." But when one is in dreamless sleep and not aware of anything, "[h]e slips out of the heart through these veins [Hitā] and rests within pericardium." Ajātaśatru then presents the following account of the vital functions and the self:

> As a spider sends forth its thread, and as tiny sparks spring forth from a fire, so indeed do all the vital functions (*prāṇa*), all the worlds, all the gods, and all beings spring from this self (*ātman*). Its hidden name (*upaniṣad*) is 'The real behind the real,' for the real consists of the vital functions, and the self is the real behind the vital functions. (BU 2.1.20)

The self (*ātman*) is said to be the source of not only the vital functions, but also of all worlds, all gods, and all beings more generally. Whereas the real is said to consist of the vital functions, it is the self alone which is taken to be, "the real behind the real." This is the first place in BU where the self plays this foundational role supporting all else, a theme that will reoccur throughout the later dialogues. It will be especially important in §4 to recall that the self is taken to be the source from which the vital functions spring.[4]

Ajātaśatru continues to teach Gārgya over the next two sections, drawing correspondences between the gods, parts of the body, and corresponding capacities. The visible appearances (*rūpa*) of *brahman* are then divided by what has fixed shape and what does not, mortal and immortal, stationary and in motion, Sat and Tyam, where the same distinctions are then applied to the body (*ātman*). Ajātaśatru goes on to present the following enigmatic passage:

[4] Although it is natural to follow Acharya (2013, p. 17) in providing a causal interpretation of the metaphors contained in the passage above, talk of what the real consists of, or of the real behind the real suggests a stronger constitutive form of explanation. See Brast-McKie (2020, Draft) for further discussion of the varieties of constitutive explanations.

> Now, the visible appearance of this person is like a golden cloth, or white wool, or a red bug, or a flame, or a white lotus, or a sudden flash of lightning. And when a man knows this, his splendor unfolds like a sudden flash of lightning. Here, then, is the rule of substitution [*ādéśá*]: "not—, not—," for there is nothing beyond this "not." And this is the name—"the real behind the real," for the real consists of the vital functions, and he is the real behind the vital functions. (BU 2.3.6)

In addition to including the first instance of *néti néti* in BU, the passage given above specifies the *ādéśá* which continues to reoccur. Whereas Olivelle (1998) adopts Thieme's (1968) translation of *ādéśá* as, "the rule of substitution," I will follow Acharya's literal interpretation of *ādéśá* as 'indication', writing:

> Literally, it is 'indication': the indicated teaching demonstrated through a discourse (object), or it is the method or means of indicating the reality (agent). In this specific case of Ajātaśatru's discourse, it is about the method: the way of apprehending and teaching the reality. (2013, p. 26)
>
> This is the method of consecutive critical negation needed to understand the complete truth of the Reality of reality. Thus, with the *ādéśá* of *néti néti* Ajātaśatru teaches Gārgya the fact that as long as one is not awakened to the totality of truth one should say 'no' to all approximated specifications or identifications as he did. (2017, p. 542)

Understanding *néti néti* as the, "method for indicating the Reality of reality," (p. 542) will help to interpret `NA` in the following section. Put roughly, I will take *néti néti* to universally quantify over properties, where all such properties fail to provide a full account of what it is to be the self (*ātman*).

In addition to defending a literal interpretation of *ādéśá* as the method of indication, Acharya (2013, p. 27) rejects Slaje's (2010, p. 28) translation (first suggested by Geldner (1908, 1928)) which takes *néti néti* to express, "an affirmative statement *via* double negation." Rather, Acharya explains that, "in a simple single sentence a proposition marked or unquoted by *íti* is negated," (p. 29) where *íti*, "[i]ndisputably means 'thus' or 'so,' and it refers to the way a certain thing is done, said, seen, thought, and so on." (p. 32) These remarks suggest that *íti* is a variable that takes predicate position, designating a *way for things to be*, but without providing an explicit description. Accordingly, *íti* ranges over a second-order domain of candidate ways for things to be.[5] By negating *íti* as in '*na íti*', or '*néti*', we have the negation of what is said about such a way for things to be, i.e., that it does not provide a full account of the self.[6] It remains, however, to determine what is intended by the repetition expressed by *néti néti* in `NA`, as well as the intended domain of values.

[5] See below for concerns about expressing higher-order quantification in English.

[6] Negation is in the first place a sentential operator, though property negation may be defined by $\neg\varphi := \lambda\vec{x}.\neg\varphi(\vec{x})$ where φ is a predicate and the open sentence '$\varphi(\vec{x})$' is negated and bound by lambda abstractors with $\vec{x}$ not occurring in φ, shifting the expression to an n-place predicate. Object negation may then be defined by $\neg a := \neg\lambda x.(a = x)$, expressing the property of not being identical to a rather than any kind of object.

In order to interpret *néti néti*, Acharya considers repeated instances of *íti* in the *Śatapatha-Brāhmaṇa* (SB) and *Mahābhārata* (MB), arguing that rather than double negation or added emphasis, the repetition expresses some degree of diversity over which the values that each *íti* may range:

> The breaths, though being thus in the middle [of it], move up and apart along the body 'in this way' and 'in that way.' (SB IX.4.3.6)
>
> Then having heard the statements of all of them, [which were stated] 'in this way' and 'in that way,' and [finally] having heard the statement of Vāsuki, Elāpatra said the following. (MB I.34.1)
>
> In both of these cases, each *íti* of the *íti céti ca* phrase stands for one statement or action, and by means of repetition, a series of such statements or actions— similar but different— is suggested. This means that in an elliptical situation *íti* can itself stand in the place of the variable implied statement or description. [. . .] Therefore, the expression of *néti néti* there negates not only two ways of definition but any 'this' or 'that' way. The pair of *néti néti* is an abbreviation for a series of consecutive negations in which each time a specification in the form of the predicate is denied, the subject in the form of the ever-existing 'It' is left unharmed. (2013, p. 33)

Not only is there some diversity among the domain of values that the variable *íti* may designate, these remarks suggest that the *íti* is bound by a universal quantifier that ranges over ways for things to be which I will refer to as *properties*.[7] Understanding this *ādéśá* will help to interpret SD in §4.

Since 'property' is a noun, there is no way to refer to a property in English without taking it to be an object. Whenever a property is the subject of an English sentence, that property is treated like an object which is then predicated. Something similar may be said about quantifying over properties since this amounts to quantifying over things that are properties. For instance, although one might wish to read the sentence 'Dor and Chelsea have something in common' as claiming that there is property common to both Dor and Chelsea, the best we can do in English is to quantify over some*thing* which is related to both Dor and Chelsea by instantiation. Nor does it help to speak of ways for things to be as I did above, since 'way' is also a noun. To put the point syntactically, instead of binding terms in predicate position, attempts to quantify over properties in English end up binding variables in nominal position. Insofar as `NA` is to be taken to claim that there is no explaining what it is to be the self where explanations relate properties not objects, English does not posses the expressive resources needed to formalise *néti néti* as it occurs in `NA`. These considerations motivate the introduction of a formal language with higher-order quantifiers that bind variables in predicate rather than nominal position. Before drawing on these resources, the following section will consider a claim made in a dialogue between Yājñavalkya and Vācaknavī. For the time being, I will continue to quantify informally over properties.

[7] See Hock (2002, p. 282) for a similar interpretation of *néti néti* in the Maitreyī Dialogue.

3 Vācaknavī Dialogue

After claiming the prize from King Janaka, Yājñavalkya is challenged by a series of Brahmins in attendance, each attempting to establish their superiority and claim the prise for themselves. Among these opponents is Gārgī Vācaknavī who asks Yājñavalkya, "tell me— since this whole world is woven back and forth on water, on what, then, is water woven back and forth?" (BU 3.6). In a series of verbal exchanges, Yājñavalkya presents the following account:

$$
\begin{array}{c}
\text{Whole World} \\
| \\
\text{Water} \\
| \\
\text{Air} \\
| \\
\text{Worlds of the Intermediate Region} \\
| \\
\text{Worlds of the Gandharvas} \\
| \\
\text{Worlds of the Sun} \\
| \\
\text{Worlds of the Moon} \\
| \\
\text{Worlds of the Stars} \\
| \\
\text{Worlds of the Gods} \\
| \\
\text{Worlds of Indra} \\
| \\
\text{Worlds of Prajāpati} \\
| \\
\text{Worlds of } \textit{brahman}
\end{array}
$$

Given any two adjacent nodes in the diagram above, Yājñavalkya explains to Vācaknavī that the higher is woven back and forth on the lower. When Vācaknavī asks, "On what, then, are the worlds of *brahman* woven back and forth?" Yājñavalkya appears to claim that there is no answer:

> Don't ask too many questions, Gārgī, or your head will shatter apart! You are asking too many questions about a deity about whom one should not ask too many questions. (BU 3.6)

Vācaknavī then fell silent, apparently satisfied with Yājñavalkya's response. However we conceive of "being woven back and forth on," it is clear that this expresses a form of explanation where the worlds of *brahman* are at the bottom of the hierarchy, explaining all else. Accordingly, I will regiment the expression 'being woven back and forth' by the primitive symbol '$\leqslant$' which I will take to express a generic form of explanation, left open to further specification.

It remains to identify the grammatical categories to which 'φ' and 'ψ' belong— i.e., their *type*— in asserting claims of the form '$\varphi \leqslant \psi$'. Since objects do not explain each other, it is important to stress that 'φ' and 'ψ' cannot take nominal position. For instance, consider the explanation that part of what it is to be gold is to have 79 protons which, or put otherwise, having 79 protons is *necessary for* being gold. Or to take another example, one might explain that intensionally harming is *sufficient for* being wrong. Put roughly, both cases provide a partial account of one property by way of another where the properties in question are expressed by predicates rather than singular terms. By contrast, it is unnatural and perhaps unintelligible to claim that one object explains another. Certainly objects are neither necessary nor sufficient for each other.[8] Instead of taking the expressions for the *explanans* and *explanandum* to be singular terms, I will assume that 'φ' and 'ψ' take predicate position in claims of the form '$\varphi \leqslant \psi$'. Accordingly, I will take 'Worlds of *brahman*' to be a predicate, where the same may be said for each expression in the diagram above. This assumption helps determine the type of 'Worlds of' as it occurs above, constraining our interpretation Yājñavalkya's account.

Whereas '*brahman*' is a name, I will take 'Worlds of *brahman*' to express the unique and all encompassing way of being *brahman* which I will refer to as *brahman*'s *haecceity*.[9] Although objects may have many properties, each object has a unique haecceity.[10] More generally, I will understand the worlds that Yājñavalkya introduces above to be the ways of being each of the entities that he considers. Accordingly, I will take Yājñavalkya's account to assert that all ways of being are woven back and forth on the way that *brahman* is— i.e., *brahman*'s haecceity. Letting 'β' refer to *brahman* and '$\lambda x.(\beta = x)$' express *brahman*'s haecceity, we may formalise this claim as follows:

(BF) $\forall X[\lambda x.(\beta = x) \leqslant X]$.

The claim above asserts that every way of being is *explained by* the way that *brahman* is. In what follows, I will also regiment `NA` by a second-order claim to the effect that no property explains what it is to be the self (*ātman*). As I will show, the forms of explanation in these two claims cannot be identified without leading to a contradiction. Nevertheless, I will argue that there is good reason to assimilate these distinct forms of explanation under a common theory of explanation. Before presenting these arguments, it will be important to begin by considering the following dialogue in which `NA` first occurs.

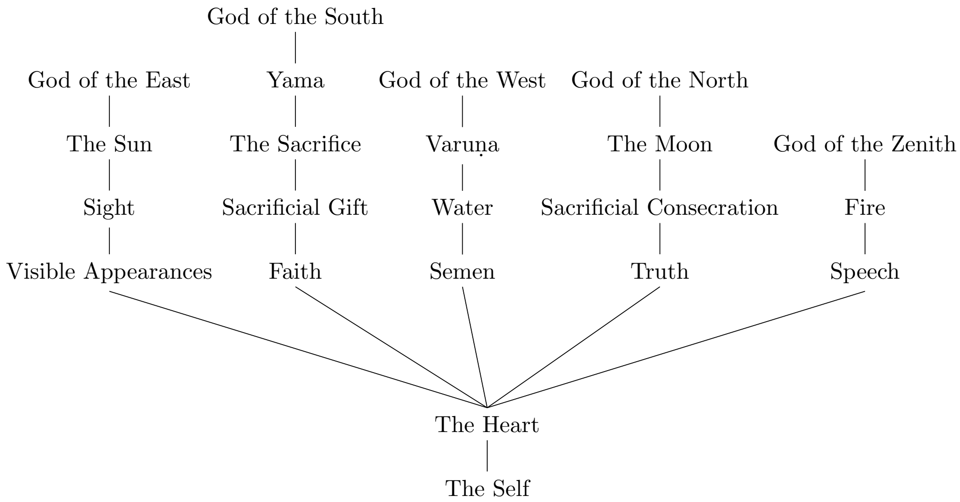
[8] It is similarly unintelligible to assert 'a because b' or 'a causes b', at least insofar as 'a' and 'b' are names that refer to objects rather than sentences that express propositions or events. Whereas 'because' and 'causes' are sentential operators, I will take 'woven back and forth on' to operate on predicates, expressing generic explanations that hold between properties rather than specific explanations that hold between propositions.

[9] The word 'haecceity' literally means *thisness*.

[10] More specifically, I will take '$\lambda x.(a = x)$' to regiment 'Worlds of a' where 'a' is a singular term. Given an open sentence φ with the free variable 'x' comparable in English to the 'it' in 'it is red', the expression '$\lambda x.\varphi$' is a one-place predicate where (*modulo* β-reduction) '$\lambda x.\varphi[\gamma]$' says of γ what φ says of x. See §5 for further discussion of lambda abstraction.

4 Śākalya Dialogue

Having withstood the interrogation of numerous opponents, Vidagdha Śākalya challenges Yājñavalkya by asking, “what is the formulation of truth (*brahman*) you know that has enabled you here to outtalk these Brahmins of Kuru and Pañcāla?” (BU 3.9.19). Yājñavalkya replies, “I know the quarters together with their gods and foundations.” This leads Śākalya to ask about each quarter, its god, and its foundation. Yājñavalkya responds by presenting the following:



In the diagram above, lines indicate foundation, where what is higher is founded on what is lower. However, as above, Yājñavalkya asserts each connection verbally, not diagrammatically. As we can see, all is founded upon the self (*ātman*), where presumable the five quarters cover everything.

Śākalya continues by asking, “On what are you and your self (*ātman*) founded?” Yājñavalkya answers by claiming that the self is founded on the out-breath, which is founded on the in-breath, which is founded on the inter-breath, which is founded on the up-breath, which is founded on the link-breath. Yājñavalkya concludes by asserting `NA`, copied below for convenience:

(`NA`) About this self (*ātman*), one can only say ‘not—, not—’.
sa eṣa neti netyātmā. (BU 3.9.26)

Given the foundation claims which Yājñavalkya makes immediately prior to `NA`, one might be tempted to take *néti néti* to indicate a series of negated foundation claims in which the self is the subject of each claim. More generally, one might take `NA` to assert that for any object whatsoever, the self is not founded on that object. However, this reading contradicts Yājñavalkya’s claims that the self is founded on the vital functions. Rather, I will take `NA` to assert that despite the role that the self plays in the theory that Yājñavalkya gives above, neither this theory nor any other provides a fully adequate account of the self. This reading echoes Acharya’s (2017, p. 542) interpretation of *néti*

néti that, "to the totality of truth one should say 'no' to all approximated specifications or identifications." Thus I will take `NA` to assert that for any way of being X, what it is to be the self (*ātman*) is not merely to X.

It remains to regiment `NA` in a language in which we may begin to study its logical relationship to other prominent claims made throughout BU. Accordingly it will be important to understand how to conceive of the candidate claims that are being successively denied. Rather than imposing a substantive theory at the outset, I will maintain a generic reading of these claims, assuming that each is the rejection of a form of explanation be it that of definition, or specification, or identification. Assuming that each '*íti*' designates a candidate explanans, the '*na*' in each '*néti*' asserts that no such candidate provides an adequate explanation of what it is to be the self (*ātman*). More generally, I will regiment `NA` as the following higher-order claim which universally quantifies over properties that do not explain what it is to be the self:

(NA) $\forall X[X \nless \lambda x.(\alpha = x)]$.

I have used 'α' in place of '*ātman*', where '$\lambda x.(\alpha = x)$' expresses the way of being *ātman*. Thus '$X < \lambda x.(\alpha = x)$' asserts that X explains what it is to be *ātman* and '$X \nless \lambda x.(\alpha = x)$' expresses its negation. By binding the variable 'X' with a second-order universal quantifier, NA assert that for any way of being, being that way does not explain what it is to be the self.

Suppose that one were to object, claiming that it is not only artificial to include the property of being identical to the self (*ātman*) or to include higher-order quantifiers, but that this cannot be what the authors of BU had in mind. This objection misconstrues the aim of formalising prominent claims made in the *Upaniṣads*, or for that matter the aim of formalising sentences of any natural language. Instead of encoding all aspects of a sentence's meaning, I will take the ambition to regiment a given claim to be the deliberate attempt to provide an abstraction, separating certain key elements of that claim's meaning from the artefacts imposed by its particular means of expression. This is not to claim, however, that formal languages do not introduce artefacts of their own. Nevertheless, formal languages are (by design) much simpler syntactically than natural languages and so facilitate systematic reasoning. In particular, the formalisation of `NA` given above is motivated by the aim to consider the logical relationships that this claim may bear to other central claims made throughout BU. Accordingly, an adequate regimentation of `NA` ought to capture the logical relationships that this claim bears to the regimentations of other prominent claims made in BU without including anything extraneous.[11]

[11] For instance, consider the following intuitively valid arguments:

(K) Kyoko is either home or at work.
Kyoko is not at work.
Therefore Kyoko is home.

(S) All men are mortal.
Socrates is a man.
Therefore Socrates is mortal.

Whereas the validity of (K) is preserved in a propositional language, (S) requires first-order resources in order to capture its validity, motivating the use of a first-order language.

5 A Logic for Explanation

In order to study the logical relationships that hold between NA and BF, let $\mathcal{L}$ be a second-order language which includes an infinite stock of names $\alpha, \beta, \gamma, \ldots$, first-order variables $x, y, z, \ldots$, n-place atomic predicates $F_n, G_n, H_n, \ldots$ for each natural number n, second-order variables $X, Y, Z, \ldots$ all taking 1-place predicate position, and logical constants for identity ($=$), abstraction (λ), weak explanation ($\leqslant$), negation ($\neg$), and conjunction ($\wedge$). I will refer to the names and first-order variables as *singular terms*, where first-order and second-order variables will be told apart by their typography. We may then provide a recursive definition of the *n-place predicates* (n-`pps`) of $\mathcal{L}$ as follows:

(F) 'φ' is an n-`pp` if φ is an n-place atomic predicate where $n \geqslant 0$.

($=$) '$a = b$' is a 0-`pp` if a and b are singular terms.[12]

(λ) '$\lambda x.\varphi$' is an $(n+1)$-`pp` if x is a first-order variable and φ is a n-`pp`.[13]

($[]$) '$\varphi[a]$' is an $(n-1)$-`pp` if φ is an n-`pp` for $n > 0$ and a is a singular term.

($\forall$) '$\forall X\varphi$' is a 0-`pp` if X is a second-order variable and φ is a 0-`pp`.

($\leqslant$) '$\varphi \leqslant \psi$' is a 0-`pp` if φ and ψ are either second-order variables or 1-`pps`.

($\neg$) '$\neg\varphi$' is a 0-`pp` if φ is a 0-`pp`.

($\wedge$) '$\varphi \wedge \psi$' is a 0-`pp` if φ and ψ are 0-`pps`.

I will refer to the 0-place atomic predicates as sentence letters, and the 0-`pps` of $\mathcal{L}$ which do not include any free variables as the *well-formed sentences* (`wfs`).[14] In particular, we may observe that BF is a `wfs` of $\mathcal{L}$.[15] It remains, however, to provide an informal interpretation of the logical constants of this language as well as a means by which to regiment NA given that '$<$' has not been included among the stock of primitive symbols that constitute $\mathcal{L}$.

Relying on an intuitive understanding of identity, negation, and conjunction for the time being, we may begin by clarifying the intended interpretation of abstraction. Using 'it' as a variable where the reference of 'it' has not already been fixed by context, anaphora, or some other means, we may consider sentences of arbitrary complexity in which 'it' occurs. For instance, just as we may say that 'it is red', we may say 'Sue wanted to buy it but John did not', and so on for sentences of even greater complexity. Given a sentence which includes

[12] I will rely on context to resolve use-mention ambiguities, avoiding corner quotes for ease.

[13] I will occasionally reuse variables as meta-variables for simplicity.

[14] If $\varphi = \lambda x.\psi$ is a n-`pp` where x occurs in ψ but λx does not occur in ψ, then x is *free* in ψ and *bound* in φ. If x is free (similarly bound) in φ, then x is also free (bound) in '$\varphi[a]$', '$\lambda y.\varphi$' for $y \neq x$, '$\forall X\varphi$', '$\neg\varphi$', '$\varphi \leqslant \psi$', '$\psi \leqslant \varphi$', '$\varphi \wedge \psi$', '$\psi \wedge \varphi$', '$\varphi \vee \psi$', and '$\psi \vee \varphi$'. If x occurs in an n-`pp` φ without being bound in φ, then x is *unbound* in φ. Although a variable may be both bound and free in a n-`pp` φ, every unbound variable in φ is free in φ.

[15] Since β is a name and x is a variable, '$\beta = x$' is a 0-`pp`, and so '$\lambda x.(\beta = x)$' is a 1-`pp`. Since X is a variable, '$\lambda x.(\beta = x) \leqslant X$' is a 0-`pp`, and so '$\forall X[\lambda x.(\beta = x) \leqslant X]$' is a 0-`pp`.

an unbound occurence of ‘it’, we may construct a corresponding predicate by prefixing the abstractor ‘is such that’ as in the predicates ‘is such that it is red’ and ‘is such that Sue wanted to buy it but John did not’. Whether we have reason to construct such predicates or not, the recipe is perfectly general: a sentence with an unbound occurence of ‘it’ may be transformed into a predicate which in combination with a singular term yields a well-formed sentence. In the symbolism above, a sentence φ in which x is the only unbound variable may be transformed into a 1-place predicate ‘$\lambda x.\varphi$’ which in application to a singular term a produces the sentence ‘$\lambda x.\varphi[a]$’. Often there is room for simplification.[16] However, independent of whether the resulting sentences can be simplified or not, their intelligibility remains.

Whereas English contains clear examples of abstraction by building complex predicates out of sentences which include unbound variables, natural languages do not include intuitive analogues of quantification into predicate position. Rather, quantifying into predicate position is part of what motivates the use of higher-order languages such as $\mathcal{L}$. Instead of relying on informal analogues, higher-order quantification is to be understood by the direct method where the syntax of $\mathcal{L}$ together with the rules of inference provided in §6 describe how to reason with higher-order quantifiers, thereby constraining their interpretation.[17] Nevertheless, it can help to conceive of quantification into predicate position as ranging over properties provided that it is understood that this is only an approximate way of thinking. In what follows, I will present arguments with the higher-order resources included in $\mathcal{L}$, postponing consideration of substitutional readings of higher-order quantification to §7.

In order to interpret the operator ‘$\leqslant$’ used to regiment BF, recall the informal claim that all worlds are woven back and forth on the world of *brahman*, where ‘woven back and forth’ expresses an explanatory relationship between ways of being. Insofar as the world of *brahman* is one such way of being, it follows that the world of *brahman* is woven back and forth on the world of *brahman*. Letting ‘$\leqslant$’ express the explanatory notion of *being woven back and forth on* and taking worlds to be ways of being where the world of *brahman* is expressed by the predicate ‘$\lambda x.(\beta = x)$’, we may formalise this implication as: $\lambda x.(\beta = x) \leqslant \lambda x.(\beta = x)$. Similarly, assuming that being *ātman* is also a way of being, it follows from BF that $\lambda x.(\beta = x) \leqslant \lambda.x(\alpha = x)$, where ‘$\lambda x.(\alpha = x)$’ expresses the way of being *ātman*. We find something different in considering the implications of NA, assuming that ‘$<$’ also expresses a form of explanation. If for any way of being, being that way does not explain what it is to be *ātman* as NA asserts, then being *ātman* is not explained by being *brahman*, i.e., $\lambda x.(\beta = x) \nless \lambda x.(\alpha = x)$. Thus we cannot identify the explantory forms expressed by ‘$\leqslant$’ and ‘$<$’ without producing a direct contradiction.

[16] If the Honda is such that Maria wanted to buy it but John did not, it follows by β-reduction (given in §6) that Maria wanted to buy the Honda but John did not.

[17] See Williamson (2003, 2013) for discussion of the direct method. It is worth observing that the logics developed by Frege and Russell were also higher-order. Although I will not do so below, one may provide a formal semantics for higher-order languages such as $\mathcal{L}$.

Although $\leqslant$ and $<$ are not coextensive, it does not follow that they are completely unrelated. Rather than taking '$<$' to be an additional primitive of the language, I will adopt the following definition in $\mathcal{L}$:

$$(<)\ \ \varphi < \psi := (\varphi \leqslant \psi) \wedge (\psi \nleqslant \varphi).$$

This definition takes expressions of the form '$\varphi < \psi$' to abbreviate expressions of the form '$(\varphi \leqslant \psi) \wedge (\psi \nleqslant \varphi)$'. Reading '$\leqslant$' as 'weakly explains' and '$<$' as 'strictly explains', the definition above states that 'φ *strictly explains* ψ' is short for 'φ *weakly explains* ψ and not *vice versa*'. Although it is convenient to adopt this convention, we need not claim that strict explanations literally abbreviate the conjunction of a weak explanation and the negation of its converse. Rather, all that is required in order to maintain $(<)$ is the following biconditional:

$$(<)'\ \ \varphi < \psi \leftrightarrow (\varphi \leqslant \psi) \wedge (\psi \nleqslant \varphi).$$

Given the necessity of the principle above, we may adopt $(<)$ as a convenient shorthand, avoiding the need to introduce further primitives into the language along with a host of interaction principles. Nevertheless, it is reasonable to doubt whether $(<)'$ holds without exception given the informal targets which '$\leqslant$' and '$<$' are intended to regiment. Rather than attempting to establish a textual basis for the principles above, I will present an abductive argument for adopting $(<)$ on the basis of the explanation it provides for NA.

Despite assuming $(<)$, it is important to observe that the interpretation of $\leqslant$ has otherwise been left open. Although one might go on to require $\leqslant$ to be reflexive in attempt to explain why $\lambda x.(\beta = x) \leqslant \lambda x.(\beta = x)$, all that follows from BF is that $\leqslant$ is not irreflexive.[18] However, instead of attempting to maintain complete neutrality from which little follows, I will make the following aditional assumption about the way that identity behaves in $\mathcal{L}$:

Sub $\varphi,\ a = b \vdash \varphi_{(b/a)}$.

I will take '$\varphi \vdash \psi$' to read 'ψ follows from φ', where a *proof* is any sequence of `wfs` which are either premises or else follow from previous `wfs` in the proof. Letting φ be a `wfs` and '$\varphi_{(b/a)}$' be the result of replacing one or more occurrences of the name a in φ with the name b, the principle above asserts that co-referring names may be freely substituted for each other *salva veratate*.[19] For example, if Kashi is south of Mount Kailāsa, then Banāras is south of Mount Kailāsa since Kashi is Banāras. However natural this inference may be, **Sub** admits of exceptions in many languages. For instance, if Janaka believed that Kashi is south of Mt. Kailāsa, it does not follow that he believed that Banāras is south of Mt. Kailāsa for Janaka may not know the city by this other name, or may simply fail to form the appropriate belief. These substitution failures can be

[18] Recall that $\leqslant$ is *reflexive* just in case $\varphi \leqslant \varphi$ for all values of φ, and *irreflexive* just in case $\varphi \nleqslant \varphi$ for all φ. Given BF, $\leqslant$ is not irreflexive, though $\leqslant$ may still fail to be reflexive.

[19] The names 'a' and 'b' *co-refer* just in case $a = b$.

explained by the *opacity* of epistiemic operators such as 'believes that' which are not only sensitive to the propositions that their arguments express, but the means by which those propositions are expressed. Accordingly, **Sub** asserts that $\mathcal{L}$ is a *transparent language* on account of excluding opaque primitives that are sensitive to differences between co-referring names.[20]

Adopting **Sub** as an assumption makes it possible to derive NA from BF since these principles include co-referring names. Although '*ātman*' and '*brahman*'— or 'α' and 'β'— differ in how they refer, what they refer to is said to be the same throughout BU. For instance, consider the following passage:

> This *brahman* is without a before and an after, without an inner and an outer. *Brahman* is this self (*ātman*) here which perceives everything. That is the teaching (BU 2.5.19)

There are many passages in BU which either assert or imply the identity of the self (*ātman*) and the Absolute (*brahman*) which I will regiment as follows:

(AB) $\alpha = \beta$.

I will take AB as well as both NA and BF to be non-logical principles which regiment prominent claims made in BU. By contrast, ($<$) and **Sub** do not occur in any guise throughout the course of BU. Rather, these assumptions provide the logical grounds by which to derive inferences that hold between such principles as AB, NA, and BF. Whereas ($<$) describes the relationship that holds between weak and strict explanation, **Sub** draws an explicit connection between NA and BF since— given AB— these principles concern different aspects of a single entity that nevertheless goes by different names.

In order to derive NA from AB and BF, I will assume a number of natural deduction rules for reasoning with conjunction, negation, and second-order universal quantification. In particular, consider the following derivation (D1):

1. $\alpha = \beta$. AB
2. $\forall X[\lambda x.(\beta = x) \leqslant X]$. BF
3. $\lambda x.(\beta = x) \leqslant F$. [2]∀E
4. $\lambda x.(\alpha = x) \leqslant F$. [1,3]**Sub**
5. $\neg[\lambda x.(\alpha = x) \nleqslant F]$. [4]DN
6. $\neg([F \leqslant \lambda x.(\alpha = x)] \wedge [\lambda x.(\alpha = x) \nleqslant F])$. [5]NC
7. $F \nless \lambda x.(\alpha = x)$. [6]($<$)
8. $\forall X[X \nless \lambda x.(\alpha = x)]$. [7]∀I

[20] In addition to preserving truth-values, one may take transparent languages to preserve the proposition expressed upon substituting co-referring terms of any type in the sentences of that language. See Brast-McKie (2021) for a propositional account of transparency.

The following section will defend the natural deduction rules employed above, arguing that D1 is *explanatory* insofar as every line of D1 may be naturally explained by previous lines with the exception of the premises. For the time being, we may observe that both ($<$) and **Sub** play an essential role in D1. Insofar as D1 provides insight into why NA holds given that AB and BF both hold, we find good reason to maintain ($<$) and **Sub**. In order to complete this argument, it remains to defend the explanatory merits of D1.

In order to gain a broader perspective of the explanation that D1 provides, it will help to begin by restating NA and BF in terms of the concepts of fundamentality and foundationality which may be defined as follows:

($\triangle$) $\triangle\varphi := \forall X(X \nless \varphi)$.

($\triangledown$) $\triangledown\varphi := \forall X(\varphi \leqslant X)$.

I will take '$\triangle\varphi$' to read 'φ is fundamental' where φ is *fundamental* just in case being φ is not strictly explained by any way of being. Additionally, '$\triangledown\varphi$' reads 'φ is foundational' where φ is *foundational* just in case every way of being is weakly explained by being φ. Thus NA asserts that being the self (*ātman*) is fundamental, and BF asserts that being the Absolute (*brahman*) is foundational. Given the natural deduction rules used in D1, we may derive the following:

Fnd $\triangledown\varphi \vdash \triangle\varphi$.

Given BF, being the Absolute (*brahman*) is foundational. Together with AB, it follows by **Sub** that being the self (*ātman*) is foundational. Thus we may conclude by **Fnd** that being the self is fundamental as claimed by NA. Insofar as each step in this argument may be shown to be explanatory, we may conclude that NA is explained by AB and BF. Put simply, being oneself is fundamental since being the Absolute is foundational and the self is the Absolute.

Even if this account does not release all of the mystery of the claims that NA, BF, and AB are intended to regiment, D1 and its restatement in terms of **Fnd** help to shed light on the roles that the self (*ātman*) and the Absolute (*brahman*) play in BU. For instance, we may observe that BF cannot be derived from NA and AB since the converse of **Fnd** does not hold. After all, there could be just two ways of being which do not weakly explain each other, making both fundamental and neither foundational. Since the converse of **Fnd** requires that $\triangledown\varphi$ holds in every possibility in which $\triangle\varphi$ holds, the converse of **Fnd** cannot be maintained.[21] Although the foundationality of the self explains the fundamentality of the self, the converse explanation does not hold. Even though BF cannot be derived from NA and AB, the *Appendix* will derive AB from BF and NA though AB is not explained by BF and NA. For the time being, the following section will argue that each of the rules employed above are explanatory, concluding that D1 is explanatory as a result.

[21] Although it is natural to require **Fnd** to be *valid* insofar as $\triangle\varphi$ is true on every interpretation of $\mathcal{L}$ in which $\triangledown\varphi$ is true, space does not suffice to present a semantic theory for $\mathcal{L}$. Rather, I will take $\varphi \vdash \psi$ to entail $\Box(\varphi \to \psi)$ so that ψ holds in any possibility in which φ holds.

6 Natural Deduction

The derivation provided in the previous section relied on a number of rules of natural deduction. Although by no means unassailable, the principles employed above are difficult to reject given the present application. Not only does the conclusion of each rule of inference follow from its premises, the truth of the conclusion may be explained by the truth of its premises. To begin with, consider the rules of inference for the second-order universal quantifier:

($\forall$E) $\forall X\varphi \vdash \varphi_{[F/X]}$ where 'F' does not occur in previous lines of the proof.

($\forall$I) $\varphi_{[F/X]} \vdash \forall X\varphi$ where 'F' does not occur in the premises of the proof.

Letting '$\varphi_{[F/X]}$' be the result of replacing all occurrences of 'X' in φ with 'F', we may observe that $\forall$E requires F to be a new predicate that does not occur on earlier lines of the proof. Although universal elimination is typically unrestricted, this latitude is not needed for the present application.

In addition to only providing the inferential powers needed for the present application, the quantifier rules given above prevent the introduction of any gaps in explanation. Assuming that 'F' does not occur in the previous lines of the proof, we may take '$\varphi_{[F/X]}$' to amount to a new notation for '$\forall X\varphi$' where 'F' expresses an arbitrary way of being rather than a particular way of being. It is important to stress that arbitrary ways of being are not particular ways of being that have the further property of being arbitrary. Rather, the use of a new predicate to indicate an arbitrary way of being may be taken to express the same kind of generality that is expressed by means of a variable bound by a universal quantifier. Stripping off the quantifier and replacing the instantial variable with a new predicate permits syntactic manipulations of the kind that quantifier-free statements enjoy while retaining the same degree of generality as the original claim. Given this understanding, I will take '$\forall X\varphi$' and '$\varphi_{[F/X]}$' to express the same claim when 'F' has been introduced as a new predicate that has not already been used to name a particular way of being. It follows that there is no explanatory gap between $\forall X\varphi$ and $\varphi_{[F/X]}$. Similarly, applications of $\forall$I do not produce any gap in explanation so long as 'F' does not express a particular way of being on account of occurring in one of the premises of the argument. Accordingly, we may freely apply the quantifier rules given above without introducing any lacunas in the resulting explanation.

Something analogous may be said for **Sub** which amounts to deriving one expression of a proposition from another expression of that same proposition. Given any sentence φ of a transparent language where $a = b$, we may substitute b for any occurrence of a in φ. Insofar as φ and $\varphi_{(b/a)}$ express the same claim whenever $a = b$, there is no opportunity for a gap in explanation to emerge on account of an application of **Sub**. Thus in addition to preserving truth, any inference that **Sub** may yield between `wfs` of $\mathcal{L}$ is guaranteed to be explanatory. Although this may not hold for languages with opaque operators, there is nothing to suggest that the principles BF, NA, and AB require $\mathcal{L}$ to include opaque operators. Rather, I will maintain the stipulation that $\mathcal{L}$ is a transparent language along with the assumption that $\mathcal{L}$ provides the expressive resources needed to regiment the Upaniṣadic claims considered above.

The next two rules of inference which occur in D1 concern the inferential behaviour of conjunction and negation. To begin with, the double negation introduction rule DN draws $\neg\neg\varphi$ as a conclusion from φ. For completeness, I will also present the corresponding elimination rule below:

(DN) $\varphi \vdash \neg\neg\varphi$.

(ND) $\neg\neg\varphi \vdash \varphi$.

Whereas ND is contentious among certain theorist, DN is much harder to dispute. For instance, suppose that one were to assume that asserting $\neg\varphi$ amounts to refraining from asserting φ rather than outright rejecting φ. On such a view of negation, asserting $\neg\neg\varphi$ amounts to refraining from asserting $\neg\varphi$, thereby leaving it open whether φ is asserted.[22] Although such a theorist may reject ND on these grounds, the same considerations do not apply to DN. Were one to assume φ, one must refrain from asserting $\neg\varphi$ given a non-assertion view of negation, from which $\neg\neg\varphi$ follows as an immediate result. Moreover, this reasoning provides a natural explanation of why $\neg\neg\varphi$ holds on a non-assertion view of negation given that φ has been granted. By contrast, one may explain on a classical view of negation that asserting φ amounts to rejecting $\neg\varphi$, and that rejecting $\neg\varphi$ amounts to asserting $\neg\neg\varphi$. Thus it is natural to maintain DN independent of which of these theories of negation one defends.

It remains to defend NC which draws the negation of a conjunction as a conclusion from the negation of either of its conjuncts. We may state this rule as follows where '$\neg\varphi/\neg\psi$' is to be replaced with either '$\neg\varphi$' or '$\neg\psi$':

(NC) $\neg\varphi/\neg\psi \vdash \neg(\varphi \wedge \psi)$.

In defence of this rule, it is worth considering what it would take for NC to fail to hold. For instance, suppose that one were to assert (G) while rejecting (A):

(G) It is not the case that Gārgya is wise.

(A) It is not the case that both Gārgya is wise and Ajātaśatru is wise.

On a classical view of negation, rejecting (A) amounts to asserting that Gārgya is wise and Ajātaśatru is wise, from which it follows that Gārgya is wise, thereby contradicting (G). Accordingly, classical views of negation and conjunction cannot maintain (G) while rejecting (A).[23] Rather, the classical theorist will accept that (A) follows from (G), appealing to (G) in order to explain why (A)

[22] I am grateful to Aditya Guntoori for pressing me on this point.

[23] Given conjunction elimination, one may appeal to an analogue of contraposition for $\vdash$ in order to derive NC: if $\varphi \vdash \psi$, then $\neg\psi \vdash \neg\varphi$. However plausible, this meta-rule is by no means on more stable ground than NC which I will take to be primitive for my purposes here.

holds. More generally, whenever $\neg\varphi$ holds, $\neg(\varphi \wedge \psi)$ holds because $\neg\varphi$ holds. Thus $\neg(\varphi \wedge \psi)$ follows from and is explained by $\neg\varphi$ (similarly $\neg\psi$).

Even in giving up a classical view of negation, there is still good reason to maintain NC. For instance, assuming a non-assertion view of negation, it follows from (G) that one cannot assert that Gārgya is wise. Accordingly, it is natural to conclude that one cannot assert Gārgya is wise and Ajātaśatru is wise, at least given the following rule for conjunction elimination:

($\wedge$E) $\varphi \wedge \psi \vdash \varphi/\psi$.

If Gārgya is wise and Ajātaśatru is wise, it follows by $\wedge$E that Gārgya is wise, thereby contradicting (G). Thus without rejecting $\wedge$E, it follows from (G) that we cannot assert that Gārgya is wise and Ajātaśatru is wise, and so (A) follows by the non-assertion theory of negation. Since nothing about this argument depends on the particular sentences substituted for φ and ψ, we may conclude more generally that NC must hold, at least insofar as $\wedge$E is to be maintained. Moreover, disputing $\wedge$E is not a plausible line to pursue since doing so is liable to impugn one's grasp of a concept bearing any resemblance to conjunction. Rather, I will take $\wedge$E to be characteristic of conjunction since this principle cannot be given up without changing the subject from conjunction to something else entirely. After all, calling a connective 'conjunction'— or else using the symbol '$\wedge$'— does not make $\wedge$ a form of conjunction. Instead, it is the inferential roles that $\wedge$ plays which characterise the concept. Although $\wedge$ may occur in principles that remain controversial, $\wedge$E is not such a principle. Given $\wedge$E, the argument above establishes that even on a non-assertion theory of negation, $\neg(\varphi \wedge \psi)$ both follows from $\neg\varphi/\neg\psi$ and is explained by $\neg\varphi/\neg\psi$. In particular, assuming $\neg\varphi$ amounts to not asserting φ which— given $\wedge$E— explains why $\varphi \wedge \psi$ cannot be asserted. Assuming a non-assertion view of negation, this latter claim amounts to asserting that $\neg(\varphi \wedge \psi)$.

Although it is possible to reject DN and NC for the sake of a particular theoretical purpose, or as the result of a particular theoretical persuasion, doing so requires theoretical intervention. By contrast, given a context in which one might reason about what follows from what prior to the imposition of exotic theoretical commitments, these rules are difficult to dispute. Put otherwise, DN and NC are rules of natural deduction insofar as they describe our pre-theoretic rational inclinations. Unless there is explicit reason to doubt their application given the theoretical commitments of the relevant context, I will presume that these rules hold without exception. In particular, BU does not provide clear grounds upon which to raise any such doubts, and so I will maintain that DN and NC not only preserve truth but are also explanatory. Having already defended the quantifier rules and provided an abductive argument for ($<$), we may conclude that all of the rules included in D1 are explanatory, even on a non-assertion view of negation. Letting a derivation be *explanatory* just in case every line of that derivation besides premises is naturally explained by previous lines, we may restate this conclusion as the claim that D1 is explanatory.

7 The Ineffable Self

The previous section argued that '$\forall X\varphi$' and '$\varphi_{[F/X]}$' may be taken to express the same general proposition so long as 'F' expresses an arbitrary way of being. Thus there cannot be any gap in explanation between $\forall X\varphi$ and $\varphi_{[F/X]}$. However, as discussed above, English does not possess the means by which to express genuine forms of higher-order quantification, binding variables in predicate position rather than in nominal position. Rather, quantification in English is achieved by generalised quantifiers which assert that entities of one kind satisfy some further condition. For instance, we may consider the role that 'all' plays in 'all men are mortal' where similar examples replace 'all' with 'some', 'no', 'most', etc. It follows that the domain of quantification is always restricted to the extension of some kind or other. We may nevertheless simulate the first-order universal quantifier in English by considering elements of a maximally inclusive kind such as expressed by 'thing', thereby asserting that every*thing* satisfies some further condition. Were one to claim to introduce a second-order universal quantifier by a similar means, generalising over ways of being rather than things, the first-orderist will object that ways of being are only a subspecies of things, not a disjoint higher-order domain of properties.[24] Instead of rejecting the claim that '$\forall X\varphi$' and '$\varphi_{[F/X]}$' express the same general proposition, the first-orderist will dispute the intelligibility of these claims as well as the intelligibility of D1, at least in its current form.

Instead of attempting to grasp higher-order quantification by the direct method, the first-orderist may seek to recover the intelligibility of D1 by taking higher-order quantification to be a restricted form of first-order quantification. One such proposal we have already found reason to reject: ways of being cannot be taken to be objects so long as '$\leqslant$' and '$<$' express explanations rather than relations between objects. However, a first-orderist need not pursue this line, maintaining a substitutional reading of higher-order quantification. Consider the following translation schema for higher-order sentences where ρ is a first-order variable restricted to the 1-place predicates of $\mathcal{L}$ and $\ulcorner\varphi_{[\rho/X]}\urcorner$ is the sentence that results from replacing all free occurrences of 'X' in φ with ρ:

(SQ) $\forall X\varphi := \forall\rho(\ulcorner\varphi_{[\rho/X]}\urcorner \text{ is true})$.

By interpreting second-order quantifiers as ranging over the 1-place predicates of $\mathcal{L}$ and not the properties which those predicates might otherwise express, the first-orderist may claim to recover the intelligibility of D1 without admitting a *sui generis* form of higher-order quantification over properties.

Given SQ, the first-orderist may reproduce an analogue of D1 assuming the truth schema that $\ulcorner\varphi\urcorner$ is true just in case φ.[25] In particular, NA is taken to be equivalent on this reading to the following first-order claim:

[24] The first-orderist's objection is based on the assumption that kinds are only of objects, and not higher-order entities such as properties, propositions, etc.

[25] Although it is natural to account for the truth of $\ulcorner\varphi\urcorner$ by appealing to φ, the same cannot be said in reverse, raising concerns that that the substitutional analogue of D1 is not explanatory.

(NA′) $\forall\rho(\ulcorner\rho \nless \lambda x.(\alpha = x)\urcorner$ is true).

Since the predicates of a language are syntactic objects, adopting SQ avoids the need for higher-order quantification. As a result, NA′ is overtly metalinguistic, quantifying over the predicates of a language in order to make claims about the truth of sentences that take a particular form. Accordingly, NA′ is limited by the expressive power of the language in question since a language with very few predicates may be unable to quantify over all of the ways of being that we might hope to express. For instance, given the reading of NA defended above, it is natural to deny that the intension was to claim that none of the predicates in $\mathcal{L}$ express ways of being which strictly explain what it is to be the self. Rather, NA sought to capture the claim that no way of being provides a strict explanation for what it is to be the self whether those ways of being are expressed by predicates in a given language or not.[26]

Instead of employing first-order quantifiers which range over the predicates of a language, a first-orderist may seek to avoid the use of quantifiers entirely by presenting a schematic regimentation of BF and NA instead. Accordingly, a theorist of this stripe may present the following proof schema (D2):

1. $\alpha = \beta.$ AB
2. $\lambda x.(\beta = x) \leqslant \varphi.$ BF*
3. $\lambda x.(\alpha = x) \leqslant \varphi.$ [1,2]**Sub***
4. $\neg[\lambda x.(\alpha = x) \nleqslant \varphi].$ [3]DN*
5. $\neg([\varphi \leqslant \lambda x.(\alpha = x)] \wedge [\lambda x.(\alpha = x) \nleqslant \varphi]).$ [4]NC*
6. $\varphi \nless \lambda x.(\alpha = x).$ [5](<)

Rather than limiting consideration to what can be expressed in a given language, we may take the sentences above to be axiom schemata which admit instances in any sufficiently expressive language by substituting the 1-place predicates in that language for φ. By taking **Sub***, DN*, and NC* to by rule schemata, providing general means by which to reason with axiom schemata, a first-orderist may claim to have derived NA* from AB and BF*.[27] It follows that for any sufficiently expressive language, given that AB and all instances of BF* hold, we may conclude by D2 that all instances of NC* hold. Thus the first-orderist may claim to have recovered the intelligibility of D1 without admitting the intelligibility of the second-order quantifiers by which D1 was stated.

It is important to observe that the interpretation of D2 given above does not quantify over all sufficiently expressive languages. Although D2 has instances in any extension of $\mathcal{L}$, the proof schema itself is not a universal claim, nor

[26] See Williamson (2013, p. 255) for related criticism of substitutional views of higher-order quantifiers, as well as Georgi (2015) general concerns about translation schemata such as SQ.

[27] We may consider AB a schema if we wish, only there are no schematic variable to replace.

are the schemata which it contains.[28] Additionally, even if it were possible to quantify over all instances of NA* across every extension of $\mathcal{L}$, it is natural to doubt that this provides a natural reading of NA. Rather than making a metalinguistic claim which quantifies over all instances of a given schema in a sufficiently expressive language, a primitive schematic reading of NA* and BF* avoids the need for quantifiers entirely. Instead of asserting general claims, each schema provides a recipe for constructing a claim that may be asserted given the resources of a sufficiently expressive language. Accordingly, NA* provides a general method for rejecting claims which attempt to strictly explain what it is to be the self by means of some other way of being. Although by no means identical, we may think of the schematic reading of NA* as providing the *ādésá* for the self which I will follow Acharya (2013) in interpreting as the method of indicating what it is to be the self. Given any attempt to strictly explain what it is to be the self, NA* specifies the negation of that claim.

Even without attempting to quantify over languages or the instances of schemata, the schematic reading of NA* has a distinctively metalinguistic flavour. Rather than asserting a propositional content, a schema can only offer a means by which to construct sentences that may then be asserted. Perhaps this is not far from what NA was intended to express: that every attempt to define, specify, or explain what it is to be the self in any language must be rejected. Although the instances across all languages of NA* may be too many to quantify over, the explicit attempts to explain what it is to be the self are finite and easy to count. Instead of attempting to reject all possible attempts to explain what it is to be the self at the outset, this schematic regimentation of NA encodes the disposition to reject such explanations as they arise in whatever language. Put otherwise, the self is ineffable, where saying so is not to ascribe the property of ineffability to the self, but to commit to rejecting any attempt to define, specify, or explain what it is to be the self as those attempts are presented. Even so, it remains possible that there is a distinct way of being that strictly explains what it is to be the self, only that we cannot express this way of being in any language. Whereas NA* leaves this possibility open, NA understood as a second-order universal claim rules this possibility out. Rather, NA asserts that the self is fundamental insofar as there is no way of being which explains what it is to be the self independent of whether those ways of being can be expressed by the predicates of a given language.

Insofar as NA is to be asserted, there is little reason to formulate NA* though such a theorist may accept its instances. By contrast, NA* cannot be asserted, but rather describes (or recommends) what is to be asserted given the resources of a sufficiently expressive language. Whereas the high-orderist may appeal to AB and BF in order to explain why it is that being the self is fundamental as claimed by NA, the analogous explanation for NA* is somewhat less satisfying. Although one may defer to D2 together with the acceptance of

[28] There are powerful set-theoretic considerations which speak against the possibility of quantifying over all languages, for quantifying over all languages puts one in a position to define new linguistic resources not already captured by the first attempt at quantification.

AB and all instances of BF*, it is natural to ask what justifies all instances of BF*. One explanation appeals to BF: it is because every way of being is weakly explained by the way that the Absolute is that all instances of BF* may be asserted. Although one could seek to understand why BF holds, BF expresses a definite proposition that one could take to be axiomatic on account of holding without further explanation. By contrast, there is no one propositional content which BF* asserts that a theorist may explain or else assume without further explanation. Were such a theorist to take BF* to be an axiom schema licensing the assertion of any of its instances, it is natural to feel dissatisfied with an account that accepts each instance but without providing any explanation common to all. Nevertheless, such an account is consistent.

Instead of attempting to rule out a schematic or substitutional reading of `NA` as incoherent, the present aim is to consider which reading is appropriate. Whereas the schematic reading leaves open the possibility of an ineffable strict explanation for what it is to be the self, appealing to a disunified variety of instances of BF* in order to account for the instances of NA*, the higher-order regimentation of `NA` avoids both of these demerits. Additionally, NA does not face problems raised by the substitutional reading of the higher-order quantifiers which either attempts the impossible by quantifying over all languages, or else restricts consideration to the ways of being that can be expressed in a single language. By contrast, NA draws on the expressive resources provided by $\mathcal{L}$ to assert that being the self is fundamental insofar as there is no distinct way of being which strictly explains what it is to be the self. Insofar as $\mathcal{L}$ is sufficient to make this universal claim, the claim itself does not thereby inherit any limitations from the language in which it is expressed. Rather, NA asserts a general claim about all ways of being, reaching well beyond what one could hope to express with the predicates included in any language.[29]

Although `NA` first occurs in the Śākalya Dialogue quoted above, `NA` reoccurs throughout BU where each occurrence is followed by the claim given below that what it is to be the self is not something that one can know:

> About this self (*ātman*), one can only say 'not—, not—.' He is ungraspable, for he cannot be grasped.[30] (BU 3.9.26)

It is natural ask why one cannot know what it is to be the self. For instance, assuming that NA* is axiomatic, one might claim that it is only by description that we may come to know what it is to be the self but there is no predicate of any language which can provides such a description. Put otherwise, being the self is unknowable be virtue of being ineffable. Here we may object that description is not the only means by which we may acquire knowledge. For instance, we may come to know people, places, and things by acquaintance, e.g.,

[29] This is not to claim that second-order quantification does not succumb to the phenomenon of indefinite extensibility. See Rayo and Uzquiano (2006) for relevant discussion.

[30] In BU 4.2.4, `NA` follows another discussion of the vital functions which are said to constitute the person. See also BU 4.4.22 and BU 4.5.15 for further occurrences of `NA`.

Janaka may know Kashi, and not merely by description. Additionally, we may come to know how to perform certain actions or skills such as how to speak, read, swim, etc. Whereas propositional knowledge is often mediated by means of description— e.g., the testimony of articulate and reliable experts— the same cannot be said for knowledge by acquaintance or know-how. For instance, there is little to be said about how to swim that could ever communicate what an expert swimmer knows to another who does not already know how to swim. Similarly, Janaka's knowledge of Kashi cannot be transmitted in words, but must be learned for oneself through the medium of one's own experience. Insofar as description does not provide the only means by which one may acquire knowledge, we cannot appeal to the ineffability of being the self to explain why one cannot know what it is to be the self.

By contrast with the schematic regimentation of `NA`, the higher-order regimentation NA provides a natural account of why one cannot grasp what it is to be the self. Instead of appealing to the ineffability of being the self to explain why being the self cannot be known, we may account for both the ineffability and the unknowability of being the self by appealing to the fundamentality of being the self: there is no way of being that can be expressed by a predicate in any language or be known by any means because there is no way of being which strictly explains what it is to be the self. Accordingly, we cannot take NA to strictly explain what it is to be the self without contradicting what NA declares. At most, NA may be taken to be part of what it is to be the self, but not all of what it is to be the self. Nevertheless, NA explains why all instances of NA* may be asserted, where this schema may be taken to formalise the *ādéśá* as the means by which to indicate what it is to be the self.

By drawing on the expressive resources of modern logic, this paper contrasts three readings of `NA`. Whereas the substitutional reading of NA was found to be parochial and overtly metalinguistic, the schematic regimentation NA* retained an metalinguistic dimension and failed to provide a unified explanation for why the instances of NA* and BF* hold. By contrast, the higher-order reading of NA avoids any language relativity, explaining why it is that being the self is fundamental by deferring to the foundationality of what it is to be the Absolute together with the identity of the self and the Absolute. Rather than claiming that these formal methods exhibit readings that were originally intended, the present ambition is to provide a regimentation that preserves the spirit of the text along the lines of the interpretation that Acharya's (2013) defends. Although higher-order languages include greater expressive resources than what is provided by English, this is no different from other formal disciplines. Rather, systematic theory is often served by the attempt to work in an artificial language where the mechanics of the primitive terms are explicitly stipulated so as to evaluate which assumptions are needed in order to present deductively valid arguments. As brought out above, very few assumptions are needed in order to derive NA from BF and AB, where the resulting derivation may be shown to be explanatory. By contrast, there are clear reasons to doubt that BF could be derived from much less explained by NA and AB.

Providing resources for discerning readings of the claims made in BU and evaluating the logical and explanatory relationships that hold between adequate regimentations of those claims help to constrain the space of interpretations that one might defend. For instance, assuming that the Vācaknavī and Śākalya dialogues are concerned to provide explanations, I showed that the forms of explanation that they employ must be distinct, ruling out interpretations which posit their identity. Nevertheless, I presented an abductive argument in support of a unified theory of explanation which accommodates the claims made in BU on minimal assumptions. Although far from conclusive, the regimentations and derivations defended above may claim to provide insight into the philosophy of the self that BU presents. In addition to demonstrating a formal methodology for extending the interpretations that have already been provided for ancient texts, the resulting regimentations and derivations may be readily integrated into analytic metaphysics. Thus the broader aim of this paper is to make the philosophy of the self presented in BU accessible to a wider contemporary audience. At least for my own sake, I am glad to be able to represent insights from the *Upaniṣads* in formal languages with which I am familiar.

Appendix

Having derived NA from BF with the help of AB and ruled out the derivation of BF from NA and AB, it remains to consider what is required to derive AB from BF and NA. We may begin by considering the following rules of inference:

($\wedge$I) $\varphi, \psi \vdash \varphi \wedge \psi$.

(CN) $\varphi, \neg(\varphi \wedge \psi) \vdash \neg\psi$.

I will take $\wedge$I to hold a similar status to $\wedge$E, where rejecting this principle amounts to a change of subject, not a non-standard theory of conjunction. Although by no means as central as $\wedge$I, we may observe that CN may be justified on both classical as well as non-assertion accounts of negation. Assuming the latter for the sake of demonstration, if φ and $\neg(\varphi \wedge \psi)$ are asserted, then $\varphi \wedge \psi$ cannot be asserted, and so ψ cannot be asserted without giving up $\wedge$I. Nevertheless, CN fails to be explanatory in the strict sense considered above since φ and $\neg(\varphi \wedge \psi)$ need not be wholly relevant to $\neg\psi$.

Whereas relatively little is required to justify the rules of inference presented above, the same cannot be said for the following definition of second-order identity $\equiv$ as well as the type-shifting principle given below:

($\equiv$) $\varphi \equiv \psi := (\varphi \leqslant \psi) \wedge (\psi \leqslant \varphi)$.

Typ $\lambda x.(a = x) \equiv \lambda x.(b = x) \vdash a = b$.

I will discuss these principles in due course. For the time being, I will draw on these assumptions in order to derive AB from BF and NA in D3 below. Although

I will argue that the resulting derivation is not explanatory, providing this derivation will nevertheless complete the picture of how NA, BF, and AB relate:

1. $\forall X[X \nless \lambda x.(\alpha = x)]$. NA
2. $\forall X[\lambda x.(\beta = x) \leqslant X]$. BF
3. $\lambda x.(\beta = x) \nless \lambda x.(\alpha = x)$. [1]∀E
4. $\neg([\lambda x.(\beta = x) \leqslant \lambda x.(\alpha = x)] \wedge [\lambda x.(\alpha = x) \nleqslant \lambda x.(\beta = x)])$. [3](<)
5. $\lambda x.(\beta = x) \leqslant \lambda x.(\alpha = x)$. [2]∀E
6. $\neg[\lambda x.(\alpha = x) \nleqslant \lambda x.(\beta = x)]$. [4,5]CN
7. $\lambda x.(\alpha = x) \leqslant \lambda x.(\beta = x)$. [6]ND
8. $[\lambda x.(\alpha = x) \leqslant \lambda x.(\beta = x)] \wedge [\lambda x.(\beta = x) \leqslant \lambda x.(\alpha = x)]$. [5,7]∧I
9. $\lambda x.(\alpha = x) \equiv \lambda x.(\beta = x)$. [8](≡)
10. $\alpha = \beta$. [10]**Typ**

We may begin by observing that ND has been used instead of DN. Although acceptable by the lights of classical logic, this principle may raise doubts for theorists committed to non-classical ways of thinking about negation. Additionally, as brought out above, one may deny that CN is explanatory even if it is assumed to preserve truth. Finally, and most importantly, we may turn to consider (≡) and **Typ** employed in the final lines of D3.

In order to evaluate (≡) it is important to reflect on what '≡' is assumed to express. Given a way for things to be φ, and a way for things to be ψ, we may assert their identity with '$\varphi \equiv \psi$'. The definition (≡) is substantive insofar as it rules out the existence of distinct properties which weakly explain each other. For instance, given this reading of '≡', it is natural to assume:

Ref $\varphi \equiv \varphi$. **Imp** $\varphi \equiv \psi, \varphi \vdash \psi$.

Sym $\varphi \equiv \psi \vdash \psi \equiv \varphi$. **Tran** $\varphi \equiv \psi, \psi \equiv \chi \vdash \varphi \equiv \chi$.

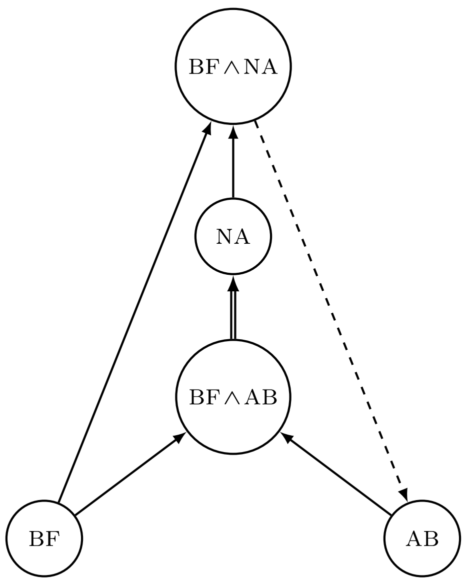
Insofar as '≡' is deserving of its name, is natural to take the principles above to hold without exception.[31] However, given (≡), each of the principles above ought to be derived from the principles provided for weak explanation instead of being taken to be axiomatic. For my purposes here, I will maintain (≡) as a convenient abbreviation that remains in need of justification. In particular, one might hope to support this convention by defending the following biconditional:

(≡)′ $(\varphi \equiv \psi) \leftrightarrow [(\varphi \leqslant \psi) \wedge (\psi \leqslant \varphi)]$.

[31] Were one to assume that $\varphi \equiv \psi \vdash \chi \equiv \chi_{(\psi/\varphi)}$ holds in $\mathcal{L}$— making $\mathcal{L}$ transparent with respect to its 1-place predicates— we may derive **Sym** and **Trans** from **Ref** and **Imp**.

Providing direct lines of support for the principle above extends beyond the scope of the present investigation, requiring additional assumptions about the role that weak explanation ought to play in developing its logic. Rather than imposing these assumptions, one may rely on an abductive argument in support of ($\equiv$) which appeals to the derivation which D3 provides. Insofar as the inference from NA and BF to AB is worth preserving, we find reason to maintain ($\equiv$) given the essential role that this definition plays in D3.

It remains to discuss **Typ** which infers the identity of a and b from the second-order identity of their haecceities. Put otherwise, **Typ** rules out the scenario where a and b share the same haecceity despite being distinct entities. Although this rule of inference is difficult to dispute, it does not follow that it is explanatory. Rather, one might assume that the second-order identity $\lambda x.(a = x) \equiv \lambda x(b = x)$ is naturally explained by $a = b$, and not *vice versa*. Nevertheless, I will assume that **Typ** preserves truth. Having defended all of the rules included in D3, we may may conclude that AB may be derived from NA and BF, though this derivation is not explanatory. Together with the explanatory derivations defended above, we may present the following picture:



Whereas the dashed arrow indicates implication, the bold arrows indicate necessary conditions. For instance, BF is necessary for the conjunction BF∧NA insofar as part of what it is for BF∧NA to hold is for BF to hold, where something similar may be said of NA. Letting φ be *inexactly sufficient for* ψ just in case there is some χ where φ is sufficient for χ and ψ is necessary for χ, I will take the double arrow to indicate inexact sufficiency. In particular, BF∧AB is inexactly sufficient for NA since BF∧AB is sufficient for NA∧AB where NA is necessary for NA∧AB. Although the implication from BF and NA to AB is not as significant as the implication from BF and AB to NA on account of failing to be explanatory, including this implication completes the present account of the relationships that hold between these principle claims of BU.

References

Acharya, Diwakar. 2013. "Néti Néti Meaning and Function of an Enigmatic Phrase in the Gārgya-Ajātaśatru Dialogue of Bṛhad Āraṇyaka Upaniṣad II.1 and II.3." *Indo-Iranian Journal* 56:3–39. ISSN 0019-7246. doi:10.1163/001972412-620402.

—. 2017. "On the Meaning and Function of Ādeśá in the Early Upaniṣads." *Journal of Indian Philosophy* 45:539–567. ISSN 0022-1791, 1573-0395. doi:10.1007/s10781-017-9321-7.

Brast-McKie, Benjamin. 2020. *Towards a Logic of Essence and Ground.* http://purl.org/dc/dcmitype/Text, University of Oxford, Oxford, UK.

—. 2021. "Identity and Aboutness." *Journal of Philosophical Logic* 50:1471–1503. ISSN 1573-0433. doi:10.1007/s10992-021-09612-w.

—. Draft. "The Construction of Possible Worlds." .

Geldner, Karl F. 1928. "Vedismus Und Brahmanismus." *Religionsgeschichtliches Lesebuch* 9.

Geldner, Karl Friedrich. 1908. *Vedismus Und Brahmanismus.* JCB Mohr.

Georgi, Geoff. 2015. "A Propositional Semantics for Substitutional Quantification." *Philosophical Studies* 172:1183–1200. ISSN 1573-0883. doi:10.1007/s11098-014-0343-7.

Hock, Hans Henrich. 2002. "The Yājñavalkya Cycle in the Bṛhad Āraṇyaka Upaniṣad." *Journal of the American Oriental Society* 122:278–286. ISSN 0003-0279. doi:10.2307/3087621.

Olivelle, Patrick. 1998. *The Early Upanishads: Annotated Text and Translation.* New York: Oxford University Press, scholar's edition. ISBN 978-0-19-512435-4.

Rayo, Agustín and Uzquiano, Gabriel (eds.). 2006. *Absolute Generality.* Clarendon Press. ISBN 978-0-19-927643-1.

Slaje, Walter. 2010. *Neti Neti: On the Meaning of an Upanisadic Citation of Some Renown in Hindu Texts and Western Minds.* Stuttgart: Franz Steiner Verlag. ISBN 978-3-515-09799-4.

Thieme, Paul. 1968. "Adesa." *Mélanges d' indianisme: à la mémoire de Louis Renou* pp.˜715–723.

Williamson, Timothy. 2003. "Everything." *Philosophical Perspectives* 17:415–465. ISSN 1520-8583. doi:10.1111/j.1520-8583.2003.00017.x.

—. 2013. *Modal Logic as Metaphysics.* Oxford: Oxford University Press.